INFORMATION AND COMPUTATION 72, 239–269 (1987)

Qualitative Robustness in Time Series

P. PAPANTONI-KAZAKOS*

University of Virginia, Thornton Hall, Charlottesville, Virginia 22901

We consider robust operations in time series. We present a definition and subsequent qualitative analysis of robustness. We also present meaningful definitions of performance criteria, such as the breakdown point and the sensitivity of robust operations. We present some specific classes of robust operations, and we discuss and analyze their properties. Finally, we fully analyze a particular class of robust predictors and interpolators, for a linearly contaminated class of stationary stochastic processes.

1. INTRODUCTION

We consider the prediction and interpolation problems, for stationary time series. The parametric formalization and solutions for those problems can be found in the pioneering works by Wiener (1949) and Kolmogorov (1941), and in several books and journal articles since then. In the present paper, we are concerned with robust formalization and solutions of the prediction and interpolation problems, where the underlying data process may be any member of some compact class of stationary processes. In contrast to the works by Masreliez and Martin (1977), and Martin and DeBow (1976), and the work by Tsaknakis and Papantoni-Kazakos (1983), we will assume that the data are observed noiselessly. Also, in contrast to the works by Tsaknakis *et al.* (1982) and Hosoya (1978), we will not restrict ourselves to linear asymptotic prediction and interpolation operations. Instead, we will define robustness as a combination of a qualitative and a game theoretic formalizations. As a result, we will propose and analyze predictors and interpolators that consist of conjunctions of both nonlinear and linear operations. The nonlinear operations are crucial for the satisfaction of qualitative robustness, where qualitative robustness for time series was first defined by Papantoni-Kazakos and Gray (1979), and it has been used for the design of robust source encoding and the design of robust quantization, by Papantoni-

* This work was supported by the Air Force Office of Scientific Research, under Grant AFOSR-78-3695.

0890-5401/87 $3.00

Kazakos (1981a, 1981b). In Papantoni-Kazakos (1984), a qualitative game theoretic formalization for robust linear filtering has been presented, and a particular conjuction of a nonlinear and a linear operation has been then proposed. In Papantoni-Kazakos (1982), performance bounds in robust filtering and smoothing have been found, when robustness is formalized as either an optimal or a suboptimal saddlepoint game. Boente *et al.* (1982) defined robustness in time series, via a contamination distance different from that in Papantoni-Kazakos and Gray (1979). As they conclude, however, the sufficient conditions derived by Papantoni-Kazakos and Gray induce strong pointwise robustness. In this paper, we will discuss the differences in the above two formalizations, and we will argue in favor of a formalization that we consider the most meaningful for the prediction and interpolation problems.

The paper is organized as follows. In Section 2 the notation and some preliminary lemmas and theorems are presented. In Section 3, the definition of robustness, as well as some additional performance criteria for robust predictors and interpolators, is presented and discussed. In Section 4, some design ideas that are consistent with our theoretical formalization are presented. In Section 5, we fully analyze a particular class of robust predictors and interpolators, for a linearly contaminated class of stationary stochastic processes.

2. Preliminaries

Let us denote by $\mathscr{F}$ some class of discrete-time, scalar stationary processes, taking values on the real line R. Let $\gamma(\cdot, \cdot)$ be some metric on R, and let $[\mu, X, R]$ denote some process in $\mathscr{F}$, where μ its measure. Let x denote an infinite data sequence, let x^l denote a given sequence of l consecutive data, and let $X_i^j; j \geqslant i$, X^l, $x_i^j; j \geqslant i$, denote respectively the sequence $X_i,..., X_j$ of random variables, a sequence of l consecutive random variables generated by the underlying process, and the sequence $x_i,..., x_j$ of observed data. Consider a sliding block window of length l with sliding step k. Let the sliding block window operate on the infinite data sequence x. Then, the above window operates sequentially on the data blocks..., x_{-k}^{l-k-1}, x_0^{l-1}, x_k^{l+k-1},... . Let g_{lk} denote a stationary operation on data blocks x_{jk}^{l+jk-1}, $j = ... -1, 0, 1,...$. If μ is some process in $\mathscr{F}$, we will denote by $\mu \circ g_{lk}$ the stationary process induced by μ and the stationary operation g_{lk}. We will denote by $\mathscr{F}_{g_{lk}}$ the class of stationary processes induced by the class $\mathscr{F}$ and the stationary operation g_{lk}. We will assume that the operation g_{lk} maps the real line R onto itself, and that the operation g_{lk} may be, in general, stochastic. By g_{l,x^l} we will then denote the measure induced by g_{lk} given the observed sequence x^l, and we will assume that the operation g_{lk} is zero memory. That is, if $g^n_{l,x}$ denotes the measure induced by the infinite sequence x on the outputs from 0 to $n-1$, we have

$$g^n_{l,x_0^{l+(n-1)k-1}}(X_0^{n-1} \in R^n) = \prod_{j=0}^{n-1} g_{l,x_{jk}^{l+jk-1}}(X_j \in R). \tag{1}$$

Given some data sequence x^n, let us form a string x of data from repetitions of x^n, and then let μ_{x^n} be the empirical measure formed by assigning probability n^{-1} on each string $T^i x$, $i=0, 1,..., n-1$, where T indicates one step shift in time. The so-formed empirical measure μ_{x^n} is (Papantoni-Kazakos and Gray, 1979) ergodic and stationary, and it induces restriction $\mu^k_{x^n}$ that are trustworthy for $k \leqslant n$. In addition to the zero memory property in (1), let the stationary operation g_{lk} also satisfy the following continuity properties:

(i) If l is given and is finite, then given $\varepsilon > 0$, given $x^l \in R^l$, $\exists \delta = \delta(l, \varepsilon, x^l) > 0$: $\gamma_l(x^l, y^l) < \delta \rightarrow \Pi_{\gamma,1}(g_{l,x^l}, g_{l,y^l}) < \varepsilon$.

(ii) Given $\mu_0 \in \mathscr{F}$, given $\varepsilon > 0$, $\eta > 0$, $\exists$ integers l_0, m, some $\delta > 0$, and for each $l > l_0$ some $\Delta^l \in R^l$ with $\mu^l_0(\Delta^l) > 1 - \eta$, such that for each $x^l \in \Delta^l$ with $\Pi_{\gamma,m}(\mu^m_{x^l}, \mu^m_{y^l}) < \delta$ it is implied that $\Pi_{\gamma,1}(g_{l,x^l}, g_{l,y^l}) < \varepsilon$. (A)

In the continuity conditions in (A), $\gamma_l(x^l, y^l) \triangleq l^{-1} \sum_i \gamma(x_i, y_i)$, and $\Pi_{\gamma,m}(\cdot, \cdot)$ is the m-dimensional Prohorov distance defined through the metric $\gamma(\cdot, \cdot)$. The definition of the Prohorov distance is in Papantoni-Kazakos and Gray (1979). Let $\rho(\cdot, \cdot)$ be some distortion measure such that, given $\varepsilon > 0$, there exist $\delta_1 > 0$, $\delta_2 > 0$, such that

$$\begin{aligned} &\rho(x, y) < \delta_1 \rightarrow \gamma(x, y) < \varepsilon \\ &\gamma(x, y) < \delta_2 \rightarrow \rho(x, y) < \varepsilon. \end{aligned} \tag{2}$$

Let $\rho_n(x^n, y^n) \triangleq n^{-1} \sum_i \rho(x_i, y_i)$. Then, the rho-bar distance $\bar{\rho}(\mu_0, \mu_1)$ between two processes μ_0, μ_1 is defined as

$$\bar{\rho}(\mu_0, \mu_1) = \sup_n \inf_{p^n \in \mathscr{P}^n} \int_{R^n \times R^n} \rho_n(X^n, Z^n)\, dp^n(X^n, Z^n), \tag{3}$$

where $\mathscr{P}^n$ is the class of all joint measures with marginals μ^n_0 and μ^n_1, and μ^n is the n-dimensional restriction of the measure μ.

From the results in Papantoni-Kazakos (1981a), we have that if the distortion measure $\rho(\cdot, \cdot)$ satisfies the conditions in (2), if the operation g_{lk} satisfies the zero memory condition in (1) and either of the conditions in (A) (depending on whether the length l of the sliding block window is finite or asymptotically long) holds, and if either $\rho(\cdot, \cdot)$ is bounded or it induces bounded values for bounded values of its arguments, and the operation g_{lk} maps the real line R onto a bounded subspace of R, then,

$$\text{Given} \quad \mu_0 \in \mathscr{F}, \quad \text{given} \quad \varepsilon > 0, \quad \exists \delta = \delta(\varepsilon, \mu_0) > 0: \quad \mu \in \mathscr{F}, \quad \bar{\rho}(\mu_0, \mu) < \delta \rightarrow \bar{\rho}(\mu \circ g_{lk}, \mu_0 \circ g_{lk}) < \varepsilon. \tag{B}$$

Property (B) is a continuity property on $\mathscr{F}$, induced by the operation g_{lk}. Furthermore, if the operation g_{lk} satisfies property (B) (or if it sufficiently satisfies properties (1) and (A)), and if $\mathscr{F}$ is a compact metric space of measures with respect to the metric $\bar{\gamma}$, then the class of measures $\mathscr{F}_{g_{lk}}$ that is induced by $\mathscr{F}$ and g_{lk} is also a compact metric space, with respect to $\bar{\gamma}$ [6]. We note that if $\gamma(\cdot, \cdot)$ is a metric on the real line, then the distance $\bar{\gamma}(\cdot, \cdot)$ in (3) (where $\rho(\cdot, \cdot)$ is substituted by $\gamma(\cdot, \cdot)$) is a metric on the space of stochastic processes.

Now let $[\mu_0, X, R]$ be a nominal stationary processes, and let the class $\mathscr{F}$ of stationary processes be defined as

$$\mu \in \mathscr{F} \leftrightarrow \bar{\gamma}(\mu_0, \mu) \leqslant \alpha, \quad \alpha \text{ a finite given constant.} \tag{C}$$

The class $\mathscr{F}$ in (C) clearly defines a convex and compact metric space, with respect to the metric $\bar{\gamma}$. Given some stationary operation g_{lk} on data realizations, given some measure μ in $\mathscr{F}$, let us select as performance criterion, or payoff function, the distance $\bar{\gamma}(\mu, \mu \circ g_{lk})$. This distance represents an error criterion applicable to a variety of problems such as the prediction and interpolation problems considered in this paper. We will present at this point some results regarding the properties of the payoff function $\bar{\gamma}(\mu, \mu \circ g_{lk})$. Our results will be presented in the form of lemmas, corollaries, and theorems.

LEMMA 1. *Let $\gamma(\cdot, \cdot)$ map bounded sets into bounded sets. Let the operation g_{lk} map the real line R onto a bounded subset of R, and let it satisfy the conditions* (A). *Let $\mathscr{M}_s$* be the class of stationary processes taking values on the real line. Then, the distance $\bar{\gamma}(\mu, \mu \circ g_{lk})$ is continuous in μ on $\mathscr{M}_s$, with respect to the metric $\bar{\gamma}$.

Proof. Let μ and μ_0 be two stationary measures. Then, clearly,

$$\begin{aligned}\bar{\gamma}(\mu, \mu \circ g_{lk}) &\leqslant \bar{\gamma}(\mu, \mu_0) + \bar{\gamma}(\mu_0, \mu \circ g_{lk}) \\ &\leqslant \bar{\gamma}(\mu, \mu_0) + \bar{\gamma}(\mu_0, \mu_0 \circ g_{lk}) + \bar{\gamma}(\mu \circ g_{lk}, \mu_0 \circ g_{lk}).\end{aligned}$$

Thus,

$$\bar{\gamma}(\mu, \mu \circ g_{lk}) - \bar{\gamma}(\mu_0, \mu_0 \circ g_{lk}) \leqslant \bar{\gamma}(\mu, \mu_0) + \bar{\gamma}(\mu \circ g_{lk}, \mu_0 \circ g_{lk}).$$

Due to the symmetric relationship, we finally have

$$|\bar{\gamma}(\mu, \mu \circ g_{lk}) - \bar{\gamma}(\mu_0, \mu_0 \circ g_{lk})| \leqslant \bar{\gamma}(\mu, \mu_0) + \bar{\gamma}(\mu \circ g_{lk}, \mu_0 \circ g_{lk}).$$

From Papantoni-Kazakos and Gray (1979) and Papantoni-Kazakos (1981a), we have that if g_{lk} satisfies the properties in the statement of the lemma, then,

Given μ_0, given $\varepsilon > 0$, $\exists \delta \triangleq \delta(\varepsilon, \mu_0) > 0$, such that

$$\bar{\gamma}(\mu_0, \mu) < \delta \rightarrow \bar{\gamma}(\mu_0 \circ g_{lk}, \mu \circ g_{lk}) < \varepsilon.$$

Thus, given μ_0, given $\varepsilon > 0$, select $\varepsilon_1 = \varepsilon/2$, $\delta = \min(\varepsilon/2, \delta(\varepsilon/2, \mu_0))$. Then,

$$\bar{\gamma}(\mu_0, \mu) < \delta \rightarrow |\bar{\gamma}(\mu_0, \mu_0 \circ g_{lk}) - \bar{\gamma}(\mu, \mu \circ g_{lk})| < \varepsilon.$$

The proof of the lemma is now complete.

In Papantoni-Kazakos (1981a), it has been found that $\bar{\gamma}(\mu_0, \mu) < \delta$ implies uniform continuity almost everywhere in R. Due to this, and directly from the results in the above paper, we can express the following corollary, which is an extension of Lemma 1.

COROLLARY 1. *Let the operation g_{lk} and the class $\mathscr{M}_s$ be as in Lemma* 1. *Let $\rho(\cdot, \cdot)$ be a distortion measure satisfying the conditions in* (2). *Then, the distance $\bar{\rho}(\mu, \mu \circ g_{lk})$ is continuous in μ on $\mathscr{M}_s$, with respect to the metric $\bar{\gamma}$.*

We will now express a lemma concerning extrema of the distance $\bar{\gamma}(\mu, \mu \circ g_{lk})$, on compact metric spaces of the measure μ. Its proof should be unnecessary.

LEMMA 2. *Let the operation g_{lk} and the class $\mathscr{M}_s$* be as in Lemma 1. Let $\mathscr{F} \subset \mathscr{M}_s$, and let it be a compact metric space, with respect to the metric $\bar{\gamma}$. Then, the distance $\bar{\gamma}(\mu, \mu \circ g_{lk})$ is bounded from above on $\mathscr{F}$, and it assumes its maximum on $\mathscr{F}$.

From the results in Papantoni-Kazakos (1981a), and as with Corollary 1, we can express the following corollary, in a straightforward fashion.

COROLLARY 2. *Let g_{lk} and $\mathscr{F}$ be as in Lemma* 2. *Let the distortion measure $\rho(\cdot, \cdot)$ satisfy conditions* (2). *Then, the distance $\bar{\gamma}(\mu, \mu \circ g_{lk})$ is bounded from above on $\mathscr{F}$, and it assumes its maximum on $\mathscr{F}$.*

LEMMA 3. *Given* μ, *the distance* $\bar{\gamma}(\mu, \mu \circ g_{lk})$ *is continuous in* g_{lk} *with respect to the metric*

$$\int_{R^L} \Pi_{\gamma,1}(g^{(1)}_{l,x^l}, g^{(2)}_{l,x^l})\, d\mu^l(x^l),$$

where $g^{(1)}_{lk}$, $g^{(2)}_{lk}$ *operations satisfy conditions* (A), *and map* R *onto a bounded subspace of* R.

Proof. We clearly have

$$|\bar{\gamma}(\mu, \mu \circ g^{(1)}_{lk}) - \bar{\gamma}(\mu, \mu \circ g^{(2)}_{lk})| \leqslant \bar{\gamma}(\mu \circ g^{(1)}_{lk}, \mu \circ g^{(2)}_{lk})$$
$$= \int_{R^l} \Pi_{\gamma,1}(g^{(1)}_{l,x^l}, g^{(2)}_{l,x^l})\, d\mu^l(x^l),$$

due to the equivalence of Π_γ and $\bar{\gamma}$ on bounded spaces. The proof is now complete.

From the results in Papantoni-Kazakos (1981a), and from Lemma 3, we can now express the following corollary.

COROLLARY 3. *Given* μ, *a class* $\mathscr{S}$ *of operations* g_{lk} *satisfying the same conditions as in Lemma* 3, *and distortion measure* $\rho(\cdot, \cdot)$ *satisfying conditions* (2), *the distance* $\bar{\rho}(\mu, \mu \circ g_{lk})$ *is continuous in* g_{lk} *with respect to the metric in Lemma* 3.

Now let $\mathscr{S}_{lk}$ denote some class of operations g_{lk}. We then express the following lemma.

LEMMA 4. *Let* $\mathscr{S}_{lk}$ *be a convex class of operations* g_{lk}, *that satisfy the conditions* (A), *and that map the real line onto a bounded subset of* R. *Let* $\rho(\cdot, \cdot)$ *be a distortion measure that satisfies conditions* (2) *and maps bounded sets into bounded sets. Let* $\mathscr{M}$ *be a convex class of stationary processes, and let for each* μ *be in* $\mathscr{M}$ *the following infimum exist*:

$$I_\rho(\mu) \triangleq \inf_{g_{lk} \in \mathscr{S}_{lk}} \bar{\rho}(\mu, \mu \circ g_{lk}).$$

Then, $I_\rho(\mu)$ is continuous in μ on $\mathscr{M}$, with respect to the metric $\bar{\gamma}(\mu_1, \mu_2)$, and it is concave in μ on $\mathscr{M}$.

Proof. Continuity of $I_\rho(\mu)$ in μ with respect to the metric $\bar{\gamma}(\mu_1, \mu_2)$ follows from the continuities of $\bar{\rho}(\mu, \mu \circ g_{lk})$, in μ and in g_{lk}, as in Corollaries 1 and 3.

Let $\mu_1, \mu_2 \in \mathscr{M}$, and let $\varepsilon: 0<\varepsilon<1$. Then, $[\varepsilon\mu_1 + (1-\varepsilon)\,\mu_2] \in \mathscr{M}$. Let g^*_{lk} be the operation in $\mathscr{S}_{lk}$ that achieves $I_\rho(\varepsilon\mu_1 + (1-\varepsilon)\,\mu_2)$. Then

$$\begin{aligned} I_\rho(\varepsilon\mu_1 + (1-\varepsilon)\,\mu_2) &= \bar{\rho}(\varepsilon\mu_1 + (1-\varepsilon)\,\mu_2, \varepsilon\mu_1 \circ g^*_{lk} + (1-\varepsilon)\,\mu_2 \circ g^*_{lk}) \\ &= \varepsilon\bar{\rho}(\mu_1, \mu_1 \circ g^*_{lk}) + (1-\varepsilon)\,\bar{\rho}(\mu_2, \mu_2 \circ g^*_{lk}) \\ &\geqslant \varepsilon I_\rho(\mu_1) + (1-\varepsilon)\,I_\rho(\mu_2). \end{aligned}$$

Thus, $I_\rho(\mu)$ is concave in μ.

We now proceed with the main theorem for this preliminary section.

THEOREM 1. *Let $\mathscr{M}$ be a convex and compact with respect to the metric $\bar{\gamma}(\mu_1, \mu_2)$ class of stationary processes. Let $\mathscr{S}_{lk}$ be a convex class of operations g_{lk} that satisfy the conditions in* (A) *and map the real line R onto a bounded subset of R. Let $\rho(\cdot,\cdot)$ be a distortion measure that satisfies the conditions in* (2), *and maps bounded sets into bounded sets. For every μ in $\mathscr{M}$, let the infimum $I_\rho(\mu)$ below exist.*

$$I_\rho(\mu) \triangleq \inf_{g_{lk} \in \mathscr{S}_{lk}} \bar{\rho}(\mu, \mu \circ g_{lk}).$$

Then, the supremum

$$\mathscr{S}_\rho(\mathscr{M}) \triangleq \sup_{\mu \in \mathscr{M}} I_\rho(\mu)$$

exists; it is unique if $I_\rho(\mu)$ is strictly concave in $\mathscr{M}$, and it is attained in $\mathscr{M}$. Furthermore, $\mathscr{S}_\rho(\mathscr{M})$ is the saddle value of a game on $\mathscr{M} \times \mathscr{S}_{lk}$, with payoff function $\bar{\rho}(\mu, \mu \circ g_{lk})$.

Proof. Due to Lemma 4, $I_\rho(\mu)$ is continuous and concave in μ on $\mathscr{M}$. Due to the continuity of $I_\rho(\mu)$, with respect to the metric $\bar{\gamma}$, and the compactness of $\mathscr{M}$, with respect to the same metric, we construct sets O_n, using $I_\rho(\mu)$. As in Lemma 2, we prove that $\mathscr{S}_\rho(\mathscr{M})$ is attained on $\mathscr{M}$. That a unique μ^* in $\mathscr{M}$ attains $\mathscr{S}_\rho(\mathscr{M})$ follows from the strict concavity of $I_\rho(\mu)$. Also, if g^*_{lk} is the operation in $\mathscr{S}_{lk}$ that attains $I_\rho(\mu^*)$, we easily obtain

$$\begin{aligned} \mathscr{S}_\rho(\mathscr{M}) = I_\rho(\mu^*) &= \sup_{\mu \in \mathscr{M}} \inf_{g_{lk} \in \mathscr{S}_{lk}} \bar{\rho}(\mu, \mu \circ g_{lk}) = \bar{\rho}(\mu^*, \mu^* \circ g^*_{lk}) \\ &\leqslant \bar{\rho}(\mu^*, \mu^* \circ g_{lk}), \qquad \forall g_{lk} \in \mathscr{S}_{lk}, \end{aligned} \tag{4}$$

$$\begin{aligned} I_\rho(\mathscr{S}_{lk}) &\triangleq \inf_{g_{lk} \in \mathscr{S}_{lk}} \sup_{\mu \in \mathscr{M}} \bar{\rho}(\mu, \mu \circ g_{lk}) = \inf_{g_{lk} \in \mathscr{S}_{lk}} \bar{\rho}(\mu_{g_{lk}}, \mu_{g_{lk}} \circ g_{lk}) \\ &\leqslant \inf_{g'_{lk} \in \mathscr{S}_{lk}} \bar{\rho}(\mu_{g_{lk}}, \mu_{g_{lk}} \circ g'_{lk}) = I_\rho(\mu_{g_{lk}}) \leqslant I_\rho(\mu^*) = \mathscr{S}(\mathscr{M}), \end{aligned} \tag{5}$$

where $\mu_{g_{lk}}$ is the process that satisfies the supremum $\sup_{\mu \in \mathscr{M}} \bar{\rho}(\mu, \mu \circ g_{lk})$. The existence of this supremum is guaranteed by Lemma 2.

But, we always have

$$\sup_{\mu \in \mathscr{M}} \inf_{g_{lk} \in \mathscr{S}_{lk}} \bar{\rho}(\mu, \mu \circ g_{lk}) \leqslant \inf_{g_{lk} \in \mathscr{S}_{lk}} \sup_{\mu \in \mathscr{M}} \bar{\rho}(\mu, \mu \circ g_{lk}). \tag{6}$$

Thus, from (5) and (6) we obtain

$$\mathscr{S}_\rho(\mathscr{M}) = I_\rho(\mathscr{S}_{lk}). \tag{7}$$

Thus, (μ^*, g^*_{lk}) is the unique solution of the saddle point game on $\mathscr{M} \times \mathscr{S}_{lk}$, with payoff function $\bar{\rho}(\mu, \mu \circ g_{lk})$. Then

$$\bar{\rho}(\mu, \mu \circ g^*_{lk}) \leqslant \bar{\rho}(\mu^*, \mu^* \circ g^*_{lk}) \leqslant \bar{\rho}(\mu^*, \mu^* \circ g_{lk}), \qquad \forall \mu \in \mathscr{M}, \quad \forall g_{lk} \in \mathscr{S}_{lk}.$$

Theorem 1 completes our preliminaries. In the next section, we define robustness for prediction and interpolation, we discuss its properties, and we present some additional performance measures for robust operations on time series.

3. Formalization of Robustness, Performance Measures

We start this section by defining robustness in time series. Then, we present sufficient conditions for its satisfaction, and we discuss the induced properties. Finally, we present and discuss some additional performance criteria for robust operations. Let $\mathscr{M}_s$ be the class of stationary processes, let $\mu_0 \in \mathscr{M}_s$ and $\mu \in \mathscr{M}_s$, let l be some given positive integer, let μ_0^l and μ^l denote the l-dimensional restrictions of the processes μ_0 and μ, respectively, let x^l and y^l denote sequences generated by the processes μ_0^l and μ^l, respectively, and let $\mathscr{P}(\mu_0^l, \mu^l)$ be the class of joint stationary measures with marginals μ_0^l and μ^l. Let $\mu^m_{x^l}$ and $\mu^m_{y^l}$ be the m-dimensional restrictions of the empirical measures determined respectively by the sequences x^l and y^l, where those empirical measures are defined in Papantoni-Kazakos and Gray (1979) and are restated in Section 2 of this paper. Then, we define

$$\Pi_{\Pi_{\gamma,m}}(\mu_0^l, \mu^l) \triangleq \inf_{p^l \in \mathscr{P}(\mu_0^l \mu^l)} \inf\{\varepsilon \colon p^l(x^l, y^l \colon \Pi_{\gamma,m}(\mu^m_{x^l}, \mu^m_{y^l}) > \varepsilon) \leqslant \varepsilon\}, \tag{7}$$

where by definition we have

$$\Pi_{\gamma,m}(\mu^m_{x^l}, \mu^m_{y^l}) = \inf\{\alpha \colon \#[i \colon \gamma_m(x_i^{i+m-1}, y_i^{i+m-1}) > \alpha] \leqslant l\alpha\}. \tag{8}$$

Thus, the distance $\Pi_{\Pi_{\gamma,m}}(\mu_0^l, \mu^l)$ is the Prohorov distance between the measures μ_0^l and μ^l, defined through the metric $\Pi_{\gamma,m}(\mu^m_{x^l}, \mu^m_{y^l})$ on data sequences.

DEFINITION 1. Given the class $\mathcal{M}_s$ of stationary processes, given μ_0 in $\mathcal{M}_s$, then,

(i) Given l and k, the operation g_{lk} is *ρ-robust at μ_0 in $\mathcal{M}_s$*, iff:

Given $\varepsilon > 0$ there exists $\delta > 0$, such that $\mu \in \mathcal{M}_s$, $\Pi_{\gamma,l}(\mu_0^l, \mu^l) < \delta \rightarrow \bar{\rho}(\mu_0 \circ g_{lk}, \mu \circ g_{lk}) < \varepsilon$.

(ii) Given k, the sequence $\{g_{lk}\}$ of operations is *ρ-robust at μ_0*, in $\mathcal{M}_s$, iff:

Given $\varepsilon > 0$, there exist integers l_0 and m, and some $\delta > 0$, such that $\mu \in \mathcal{M}_s$, $\Pi_{\Pi_{\bar{\gamma},m}}(\mu_0^l, \mu^l) < \delta \rightarrow \bar{\rho}(\mu_0 \circ g_{lk}, \mu \circ g_{lk}) < \varepsilon$, $\forall l > l_0$.

DEFINITION 2. Given the class $\mathcal{M}_s$ of stationary processes, given μ_0 in $\mathcal{M}_s$, then,

(i) Given l and k, the operation g_{lk} is *weakly ρ-robust at μ_0* in $\mathcal{M}_s$, iff:

Given $\varepsilon > 0$, there exists $\delta > 0$, such that $\mu \in \mathcal{M}_s$, $\bar{\gamma}(\mu_0, \mu) < \delta \rightarrow \bar{\rho}(\mu_0 \circ g_{lk}, \mu \circ g_{lk}) < \varepsilon$.

(ii) Given k, the sequence $\{g_{lk}\}$ of operations is *weakly ρ-robust at μ_0* in $\mathcal{M}_s$, iff:

Given $\varepsilon > 0$, there exists integer l_0, and some $\delta > 0$, such that $\mu \in \mathcal{M}_s$, $\bar{\gamma}(\mu_0, \mu) < \delta \rightarrow \bar{\rho}(\mu_0 \circ g_{lk}, \mu \circ g_{lk}) < \varepsilon$, $\forall l > l_0$.

From Definitions 1 and 2 above, we note that since $\bar{\rho}$ closeness implies Prohorov closeness; ρ-robustness at μ_0 implies weak ρ-robustness at μ_0. We also note that in both Definitions 1 and 2, the rho-bar distance is used as a stability criterion. This is in contrast to Papantoni-Kazakos and Gray (1979), where the stability criterion used is the Prohorov distance. The latter imposes milder conditions on the robust operations than the former. However, we consider the rho-bar stability criterion more appropriate, especially in time-series-oriented problems such as the prediction interpolation and filtering ones. Regarding the contamination criterion, in Definition 2, the $\bar{\gamma}$ distance is used, as in Papantoni-Kazakos and Gray (1979). On the other hand, the contamination criteria used in Definition 1 are Prohorov distances, where the Prohorov distance $\Pi_{\Pi_{\bar{\gamma},m}}$ has also been used by Boente *et al.* (1982). We note that operations g_{lk} that utilize a block window of finite length l induce processes $\mu \circ g_{lk}$ whose characteristics are determined only by the l-dimensional restriction μ^l of the measure μ. Thus, the Prohorov distance $\Pi_{\gamma,l}(\mu_0^l, \mu^l)$ is then sufficient as a contamination measure. The Prohorov distance $\Pi_{\Pi_{\bar{\gamma},m}}$ used as a contamination measure, when a sequence, $\{g_{lk}\}$, of operations is considered, basically measures contaminations on data sequences. As pointed out in Boente *et al.* (1982), contamination on data sequences is probably the most appropriate model when robustness is concerned. We point out here that in Boente *et al.* (1982), the supremum $\sup_l \Pi_{\Pi_{l,m}}(\mu_0^l, \mu^l)$ has been used as a contamination measure in robustness. Since the existence of this supremum and the value, l', at which it is attained (given that it exists) are not guaranteed, our definition of ρ-robustness (Definition 1(ii)) is different from that offered by Boente *et al.* (1982). As pointed out by Cox (1978), and discussed by Boente *et al.* (1982), linear operations may be weakly ρ-robust (Definition 2). Such operations are not, however, ρ-robust, in general. We now proceed with a theorem.

THEOREM 2. *Given the class $\mathcal{M}_s$ of stationary processes, given μ_0 in $\mathcal{M}_s$, then,*

(i) *Given l and k, let the operation g_{lk} satisfy part* (i) *of condition* (A). *Let $\rho(\cdot, \cdot)$ satisfy the conditions in* (2) *and either let it be bounded or let it map bounded sets into bounded sets, and then let g_{lk} map the real line R on a bounded subset of R. Then, g_{lk} is ρ-robust at μ_0 in $\mathcal{M}_s$.*

(ii) *Given k, let the sequence $\{g_{lk}\}$ of operations satisfy part* (ii) *of condition* (A). *Let $\rho(\cdot, \cdot)$ satisfy the conditions in* (2) *and either let it be bounded or let it map bounded sets into bounded sets, and then let g_{lk} map the real line R on a bounded subset of R. Then, $\{g_{lk}\}$ is ρ-robust at μ_0 in $\mathcal{M}_s$.*

The proof of the theorem can be basically found in Papantoni-Kazakos and Gray (1979) and Papantoni-Kazakos (1981a). Clearly, if the conditions in Theorem 2 are satisfied, the corresponding operations are weakly ρ-robust at μ_0, as well. We will now present two definitions concerning robustness within classes of measures.

DEFINITION 3. Given a convex class $\mathcal{M}$ of stationary processes that is also compact with respect to the metric $\bar{\gamma}$, then,

(i) Given l and k, given a class $\mathcal{S}_{lk}$ of operations, the operation g_{lk}^* in $\mathcal{S}_{lk}$ is *ρ-robust on $\mathcal{M} \times \mathcal{S}_{lk}$*, iff: g_{lk}^* is ρ-robust at every μ in $\mathcal{M}$, and there exists μ^* in $\mathcal{M}$ such that

$$\bar{\rho}(\mu, \mu \circ g_{lk}^*) \leqslant \bar{\rho}(\mu^*, \mu^* \circ g_{lk}^*) \leqslant \bar{\rho}(\mu^*, \mu^* \circ g_{lk}), \qquad \forall \mu \in \mathcal{M}, \ \forall g_{lk} \in \mathcal{S}_{lk}.$$

(ii) Given k, given a class $\mathcal{S}$ of sequences, $\{g_{lk}\}$, of operations, the sequence $\{g_{lk}^*\}$ in $\mathcal{S}$ is *ρ-robust on $\mathcal{M} \times \mathcal{S}$*, iff: $\{g_{lk}^*\}$ is ρ-robust at every μ in $\mathcal{M}$, and there exist μ^* in $\mathcal{M}$ and integer l_0, such that

$$\bar{\rho}(\mu, \mu \circ g_{lk}^*) \leqslant \bar{\rho}(\mu^*, \mu^* \circ g_{lk}^*) \leqslant \bar{\rho}(\mu^*, \mu^* \circ g_{lk}),$$
$$\forall \mu \in \mathcal{M}, \ \forall g_{lk} \in \mathcal{S}_{lk}, \ \forall l > l_0.$$

DEFINITION 4. Given a convex class $\mathcal{M}$ of stationary processes and classes $\mathcal{S}_{lk}$ and $\mathcal{S}$ of operations, as in Definition 3, the operation g_{lk}^* and the sequence $\{g_{lk}^*\}$ are *weakly ρ-robust* on $\mathcal{M} \times \mathcal{S}_{lk}$ and $\mathcal{M} \times \mathcal{S}$, respectively, if the parallel to Definition 3 conditions hold, where ρ-robustness at every μ in $\mathcal{M}$ is substituted by weak ρ-robustness at every μ in $\mathcal{M}$.

The following theorem connects the analysis in Section 2 and the present section, with ρ-robustness in Definition 3.

THEOREM 3. *Let $\mathcal{M}$, $\mathcal{S}_{lk}$, and $\mathcal{S}$ be as in Definition 3. In addition, let $\mathcal{S}_{lk}$ be a convex class of operations that satisfy part (i) of condition (A), and let $\mathcal{S}$ be a convex class of sequences of operations that satisfy part (ii) of condition (A), at every μ in M. Also, let $\rho(\cdot,\cdot)$ satisfy conditions (2), and either let it be bounded or let it map bounded sets into bounded sets while the operations in $\mathcal{S}_{lk}$ and $\mathcal{S}$ also map the real line, R, on a bounded subset of R. Let the following infima exist:*

$$I_{\rho,l}(\mu) \triangleq \inf_{g_{lk} \in \mathcal{S}_{lk}} \bar{\rho}(\mu, \mu \circ g_{lk})$$

$$I_{\rho,l_0,l}(\mu) \triangleq \inf_{\{g_{lk}\} \in \mathcal{S}} \bar{\rho}(\mu, \mu \circ g_{lk}), \qquad \forall l > l_0, \text{for some } l_0.$$

Then, a unique ρ-robust on $\mathcal{M} \times \mathcal{S}_{lk}$ operation, g_{lk}^*, and a unique ρ-robust on $\mathcal{M} \times \mathcal{S}$ sequence of operations, $\{g_{lk}^*\}$, exist. They are such that

$$\begin{aligned} I_{\rho,l}(\mu^*) &\triangleq \sup_{\mu \in \mathcal{M}} I_{\rho,l}(\mu) = \bar{\rho}(\mu^*, \mu^* \circ g_{lk}^*) \\ &= \inf_{g_{lk} \in \mathcal{S}_{lk}} \bar{\rho}(\mu^*, \mu^* \circ g_{lk}) \\ I_{\rho,l_0,l}(\mu^*) &\triangleq \sup_{\mu \in \mathcal{M}} I_{\rho,l_0,l}(\mu) = \bar{\rho}(\mu^*, \mu^* \circ g_{lk}^*) \\ &= \inf_{\{g_{lk}\} \in \mathcal{S}} \bar{\rho}(\mu^*, \mu^* \circ g_{lk}), \qquad \forall l > l_0. \end{aligned}$$

The proof of the theorem is straightforward from Theorems 1 and 2 and Definitions 1 and 3. We note that for the satisfaction of weak ρ-robustness on $\mathcal{M} \times \mathcal{S}_{lk}$ and $\mathcal{M} \times \mathcal{S}$, respectively, the requirements for $\rho(\cdot,\cdot)$ boundness or g_{lk} boundness for the satisfaction of condition (A) may be eliminated. As we pointed out earlier, under mild conditions, linear operations and unbounded $\rho(\cdot,\cdot)$ may provide a weakly ρ-robust solution, either on $\mathcal{M} \times \mathcal{S}_{lk}$ or on $\mathcal{M} \times \mathcal{S}$. For example, let $\mathcal{M}$ include the Gaussian process, and let the spectral densities of the process in $\mathcal{M}$ form a compact class. Let $\mathcal{S}$ be the class of one-step prediction operations. Then, the Gaussian process with the "flattest" spectral density within the compact class of spectral densities and its optimal linear predictor provide the weakly ρ-robust on $\mathcal{M} \times \mathcal{S}$ solution, if $\rho(u, v) \triangleq (u-v)^2$. The above linear predictor is not ρ-robust at any μ in $\mathcal{M}$, however. This is so because this predictor is not bounded, and it does not satisfy condition (A).

We now proceed by presenting some performance measures for robust operations. In particular, we define the breakdown point and the sensitivity of a robust operation. Then, we discuss and analyze those two performance measures. We first present two definitions, distinguishing between ρ-robust and weakly ρ-robust operations.

DEFINITION 5. Given the class, $\mathcal{M}_s$, of stationary processes, given μ_0 in $\mathcal{M}_s$, given a sequence $\{g_{lk}\}$ of operations that is ρ-robust at μ_0 in $\mathcal{M}_s$, then,

(i) The *breakdown point of the sequence* $\{g_{lk}\}$ *at* μ_0 in $\mathcal{M}_s$ is this constant α^* such that, for every $\alpha > \alpha^*$, there exist integers $m = m(\alpha)$, $l_0 = l_0(\alpha)$, such that for $\mathcal{M}_l(\alpha) = \{\mu \in \mathcal{M}_s, \mu^l \colon \Pi_{\Pi_{\gamma,m}}(\mu_0^l, \mu^l) < \alpha\}$, the supremum $\sup_{\mu \in \mathcal{M}_l(\alpha)} \bar{\rho}(\mu, \mu \circ g_{lk})$ is independent of α, $\forall l > l_0$.

(ii) The *sensitivity*, $\mathcal{S}_\rho(\mu_0, \{g_{lk}\})$, *of the sequence* $\{g_{lk}\}$ *at* μ_0 *in* $\mathcal{M}_s$, is defined as

$$S_\rho(\mu_0, \{g_{lk}\}) \triangleq \lim_{\lim_{l\to\infty} \Pi_{\Pi_{\gamma,m}}(\mu_0^l, \mu^l) \to 0} \; \lim_{l\to\infty} \frac{\bar{\rho}(\mu_0 \circ g_{lk}, \mu \circ g_{lk})}{\Pi_{\Pi_{\gamma,m}}(\mu_0^l, \mu^l)},$$

where the integer m is as in Definition 1.

DEFINITION 6. Given the class, $\mathcal{M}_s$, of stationary processes, given μ_0 in $\mathcal{M}_s$, given a sequence $\{g_{lk}\}$ of operations that is weakly ρ-robust at μ_0 in $\mathcal{M}_s$, then

(i) The *breakdown point of the sequence* $\{g_{lk}\}$ *at* μ_0 in $\mathcal{M}_s$, is this constant α^* such that, for every $\alpha > \alpha^*$, there exists integer $l_0 = l_0(\alpha)$, such that, for $\mathcal{M}(\alpha) < \{\mu \in \mathcal{M}_s, \bar{\gamma}(\mu_0, \mu) < \alpha\}$, the supremum $\sup_{\mu \in \mathcal{M}(\alpha)} \bar{\rho}(\mu, \mu \circ g_{lk})$ is independent of α, $\forall l > l_0$.

(ii) The *sensitivity*, $\mathcal{S}_{w\rho}\{g_{lk}\})$, *of the sequence* $\{g_{lk}\}$ *at* μ_0 *in* $\mathcal{M}_s$, is defined as

$$S_{w\rho}(\mu, \{g_{lk}\}) = \lim_{\bar{\gamma}(\mu_0, \mu) \to 0} \; \lim_{l\to\infty} \frac{\bar{\rho}(\mu_0 \circ g_{lk}, \mu \circ g_{lk})}{\bar{\rho}(\mu_0, \mu)}.$$

From Definitions 5 and 6, we note that the breakdown point and the sensitivity have been defined only asymptotically; that is, for operations that utilize asymptotically large sliding-block windows. We also note that bounded sensitivity corresponds to the concept of differentiability of real functions, while ρ-robustness or weak ρ-robustness corresponds instead to local continuity of real functions. We want to point the attention of the reader to our definition of the breakdown point. In contrast to Hampel's (1971) definition, we define the breakdown point at this level of contamination above which induced error is not controlled by the contamination itself. Hampel defined the breakdown point as this level of contamination above which the induced error becomes infinity. The latter definition is too restrictive for our models, in this paper. Indeed, since our performance measure in robustness is the rho-bar distance, rather than the Prohorov distance, our robust operations will be, in general, bounded. Thus, the error induced by such operations will be then bounded for every input process and, therefore, for all contamination levels. We point out here that the breakdown curve found by Tsaknalis *et al.* (1983) is consistent with Definitions 5 and 6. The linear filtering operations in Tsaknakis *et al.* (1983) are, in addition, weakly ρ-robust, in general. Finally, we note that under mild conditions, linear operations may induce bounded sensitivity, $S_{w\rho}(\cdot, \cdot)$ (previously, we pointed out that such linear operations may be weakly ρ-robust). Such operations will not, however, induce bounded sensitivity $S_\rho(\cdot, \cdot)$, since, as pointed out previously, they are not even ρ-robust. We will quantify some of the remarks made in this section in a later section, where some specific operations will be discussed. In the next section, we will discuss some design ideas that are consistent with the theory we have presented to this point.

4. Design Ideas

In this section, we will present some ideas, regarding the design of robust operations for time series, where robustness is consistent with Definitions 1 and 3, in Section 3. From the analysis presented in the previous section, we conclude that for ρ-robustness it is sufficient to consider operations that satisfy condition (A) and map the real line, R, onto a bounded subspace of R, and to adopt a distortion measure $\rho(\cdot, \cdot)$ that satisfies the conditions in (2). For the metric $\gamma(x, y) \triangleq |x - y|$ on the real line, such a distortion measure $\rho(\cdot, \cdot)$ is given by $\rho(x, y) \triangleq (x - y)^r$, where r is some positive integer. Regarding the choice of appropriate sufficient classes of operations, either one of the following classes may be considered:

$$\mathscr{S}_l^1 = \{f(x^l): \text{ scalar, real, bounded, deterministic, and continuous with respect to the metric } \gamma_l(x^l, y^l) = l^{-1} \textstyle\sum_{i=1}^l |x_i - y_i| \text{ function, where } l \text{ is a fixed, finite, and positive integer.}\} \tag{D}$$

$\mathscr{S}_l^2 = \{ g_{l,x^l}$: Scalar, stochastic, memoryless, and stationary channel, with outputs taking values on a bounded subspace of R, that is also continuous, in the sense of part (i), conditions (A). Here, l is a fixed, finite, and positive integer.$\}$ (E)

$\mathscr{S}(\mu) = \{ \{ g_{l,x^l} \}$: a sequence of stochastic, scalar, and memoryless stationary channels, with outputs taking values on a bounded subspace of R, that also satisfy part (ii), conditions (A), at the given stationary process μ.$\}$ (F)

We note that the class $\mathscr{S}_l^1$ in (D) satisfies part (i) in conditions (A), where the Prohorov distance $\Pi_{y,1}(g_{l,x^l}, g_{l,y^l})$ becomes then equal to $|f(x^l) - f(y^l)|$. Below, we state two subclasses, $\mathscr{H}_l^1$ and $\mathscr{H}_l^2$, of the class $\mathscr{S}_l^2$ in (E), that are convenient for design considerations:

$\mathscr{H}_l^1 = \{ g_{l,x^l} = f(x^l) + v$: $f(x^l)$ belongs to class $\mathscr{S}_l^1$ in (D), and v is an absolutely continuous random variable that takes values on a bounded interval of the real line, and has continuous density function.$\}$ (G)

$\mathscr{H}_l^2 = \{ g_{l,x^l} = g(h(x^l) + v)$: $h(x^l)$ is a scalar, real, deterministic, and continuous with respect to the metric $l^{-1} \sum_{i=1}^{l} |x_i - y_i|$ function. v is some absolutely continuous random variable that takes values on the real line, and has continuous density function. $g(\cdot)$ is a deterministic, scalar, real, and bounded function.$\}$ (H)

Let $\mathscr{C}$ be the class of real, scalar, and bounded functions, that also take a finite number of values. Then, a subclass of the class $\mathscr{H}_l^2$ in (H) is determined by $g(\cdot)$ functions in $\mathscr{C}$. We will denote this subclass $\mathscr{H}_l(\mathscr{C})$. The class $\mathscr{H}_l(\mathscr{C})$ of operations is of particular interest. Indeed, given some stationary process μ, this class induces a sequence $\{ g_{l,x^l} \}$ of operations that is contained in the class $\mathscr{S}(\mu)$ in (F). To see that specifically, let us define the following subclass $\mathscr{H}(\mu)$, contained in $\mathscr{S}(\mu)$:

$\mathscr{H}(\mu) = \{ \{ g_{l,x^l} \}$: Given finite positive integers l_1 and k, such that $1 \leqslant k \leqslant l_1$,

$$g_{l,x_0^{l-1}} = \{ \lfloor k^{-1}(l - l_1) \rfloor \}^{-1} \sum_{j=0}^{\lfloor k^{-1}(l-l_1) \rfloor} g(h(x_{jk}^{l_1+jk-1}) + v), \qquad \forall l \geqslant l_1, \quad \text{(I)}$$

where $\lfloor\ \rfloor$ means integer part, $h(x^l)$ and v are as in class $\mathscr{H}_l^2$, and $g \in \mathscr{C}$.$\}$

The operations g_{l,x^l} in class $\mathscr{H}(\mu)$ satisfy part (ii) of conditions (A). This can be easily verified. Indeed, the operation $g(h(x_{jk}^{l_1+jk-1})+v)$ belongs to the class $\mathscr{H}_l^2$ in (H); thus, given μ, it induces by means of μ a discrete, finite valued stationary process $\mu \circ g$ that satisfies part (i) in Definition 1. Furthermore, the averaging on the operations $g(h(x_{jk}^{l_1+jk-1})+v)$, in $g_{l,x_0^{l-1}}$, satisfies part (ii) of conditions (A) when applied to the process $\mu \circ g$. We finally observe that the class $\mathscr{H}(\mu)$ in (I) is basically independent of the particular process μ. Thus, this class is contained in the class $\mathscr{S}(\mu)$ in (F), for all stationary process μ. It can be shown that the class $\mathscr{H}(\mu)$ can be generalized to a class $\mathscr{H}_L$, such that

$$g_{l,x_0^{l-1}} = \sum_{j=0}^{\lfloor k^{-1}(l-l_1)\rfloor} a_{lj}\, g(h(x_{jk}^{l_1+jk-1})+v), \qquad \forall l \geq l_1,$$

where g, h, v are as in class $\mathscr{M}(\mu)$ and $a_{lj} > 0$, $\forall_j$, $\forall l$, (9)

$$\sum_{j=0}^{\lfloor k^{-1}(l-l_1)\rfloor} a_{lj} = 1, \qquad \forall l$$

$$\lim_{l\to\infty} \sum_{j=0}^{\lfloor k^{-1}(l-l_1)\rfloor} a_{lj}^2 = 0.$$

Indeed, the summation in (9) then satisfies part (ii) in conditions (A).

We note that the classes $\mathscr{S}_l^1$, $\mathscr{H}_l^1$, $\mathscr{H}_l^2$, $\mathscr{H}(\mu)$, and $\mathscr{H}_L$ that we presented and discussed above were all tailored toward the satisfaction of ρ-robustness in Definition 1. Furthermore, all these classes involve bounded operations on data sequences x that are generated by the acting stationary process μ. If the distortion measure $\rho(\cdot,\cdot)$, used for performance evaluation, is not bounded, then such operations may induce high errors, at processes μ, whose support is noncompact. This may happen, independently, whether or not ρ-robustness (Definition 1) is satisfied. To eliminate this problem, we may use a one-to-one bounded transformation on the data of the underlying process. Such a transformation is induced, for example, by a bounded, real, deterministic, scalar, and strictly monotone function $G(\cdot)$. If g_{l,x^l} is an operation in any one of the classes $\mathscr{S}_l^1$, $\mathscr{H}_l^1$, $\mathscr{H}_l^2$, $\mathscr{H}(\mu)$, and $\mathscr{H}_L$, we may then assume that this operation maps the variable $G(X_0)$, where X_0 is a random variable from the acting stationary process μ. Thus, given a function $G(\cdot)$, as above, let us define the class $\mathscr{M}(G)$ of stationary processes as

$$\mathscr{M}(G) = \{v = \mu \circ G \colon \mu \in \mathscr{M}\}. \tag{J}$$

If the $G(\cdot)$ function is as explained above, it represents a stationary operation; thus every process v in $\mathscr{M}(G)$ is stationary. Also, since $G(\cdot)$ is bounded, if $\mathscr{M}$ is a compact class of stationary processes, so is the class

$\mathscr{M}(G)$. Finally, given the value u of the random variable X_0, generated by the process μ, the value $G(\mu)$ is unique. Inversely, given the value v of the transformation $G(u)$, the value u is also unique.

Due to the preceding discussion in this section, given a compact class $\mathscr{M}$ of stationary processes and given $\rho(\cdot, \cdot)$ distortion measure that satisfies the conditions (2) in Section 2 for ρ-robustness, it is now sufficient to consider saddle-point games (Definition 3) on $\mathscr{M}(G) \times \mathscr{S}$, where $\mathscr{M}(G)$ is as in ((J), $G(\cdot)$ is some real, scalar, deterministic, bounded, and strictly monotone function, and $\mathscr{S}$ is any one of the classes $\mathscr{S}_l^1$, $\mathscr{H}_l^1$, $\mathscr{H}_l^2$, $\mathscr{H}(\mu)$, and $\mathscr{H}_L$. We now proceed with a lemma whose proof is obvious.

LEMMA 5. *Let $G(\cdot)$ be a given real, deterministic, scalar, continuous, bounded, and strictly monotone function. Let x^l denote some data sequence of length l that does not include x_0. Let μ be some stationary process, such that $E_\mu\{G(X_0) \mid x^l\}$ is continuous in x^l, with respect to the metric $l^{-1} \sum_{i=1}^{l} |x_i - y_i|$. Consider the infimum*

$$I_{G,l}(\mu, \mathscr{S}) \triangleq \inf_{g_{l,x^l} \in \mathscr{S}} E_\mu\{G(X_0) - g_{l,x^l}\}^2. \tag{10}$$

Then,

(i) *The infimum $I_{G,l}(\mu, \mathscr{S}_l^1)$ exists, is unique, and is satisfied by*

$$f(x^l) = E_\mu\{G(X_0) \mid x^l\}$$

and

$$I_{G,l}(\mu, \mathscr{S}_l^1) = E_\mu\{G(X_0) - E_\mu\{G(X_0) \mid x^l\}\}^2.$$

(ii) *If the class $\mathscr{H}_l^1$ in (G) is such that the random variable v is zero mean, has variance $\sigma_v^2 < \infty$, and is independent of the process μ, then $I_{G,l}(\mu, \mathscr{H}_l^1)$ exists, is unique, and is satisfied by*

$$f(x^l) = E_\mu\{G(X_0) \mid x^l\}$$

and

$$I_{G,l}(\mu, \mathscr{H}_l^1) = \sigma_v^2 + E_\mu\{G(X_0) - E_\mu\{G(X_0) \mid x^l\}\}^2.$$

We note that if μ is a stationary process whose density functions f_μ^n exist, for every dimensionality n, and if $\forall n$, $f_\mu^n(x^n)$ is compact and continuous in x^n, with respect to the metric $n^{-1} \sum_{i=1}^{n} \gamma(x_i, y_i)$, then $\forall l$, the expected value $E_\mu\{G(X_0) \mid x^l\}$ is continuous in x^l with respect to the same metric as above.

Let us now consider the class $\mathscr{H}_l^2$ in (H). Let us denote by $\mathscr{H}_l^2(\text{sgn})$ the subclass of $\mathscr{H}_l^2$, determined by $g(x) = \text{sgn}\, x = \{0; x \leqslant 0 \text{ or } 1; x > 0\}$. Let the random variable v in $\mathscr{H}_l^2(\mathrm{sgn})$ have a distribution function denoted by F_v. Let F_v be such that $F_v(-x) = 1 - F_v(x)$. Then, for the given function $h(x^l)$ in $\mathscr{H}_l^2(\mathrm{sgn})$, we obtain

$$E\{\mathrm{sgn}(h(x^l)+v) \mid x^l\} = E^2\{\mathrm{sgn}(h(x^l)+v) \mid x^l\} = F_v(h(x^l)). \tag{11}$$

Given some function $G(\cdot)$, as in Lemma 5, given a stationary process μ, we also then obtain

$$\begin{aligned} &E_\mu\{G(X_0) - \mathrm{sgn}(h(x^l)+v)\}^2 \\ &\quad = E_\mu\{E_\mu\{G^2(X_0) \mid x^l\} - 2F_v(h(x^l))\, E_\mu\{G(X_0) \mid x^l\} + F_v(h(x^l))\} \\ &\quad = E_\mu\{G(X_0) - F_v(h(x^l))\}^2 + E_\mu\{F_v(h(x^l)) - F_v^2(h(x^l))\}. \end{aligned} \tag{12}$$

From expression (12), for the class $\mathscr{H}_l^2(\mathrm{sgn})$, the following ideas evolve naturally. Given an infinite sequence x from some real and scalar stationary process μ, consider one-step sliding block window of length m that operates sequentially on the data blocks, $\ldots, x_{-m}^{1}, x^0_{m+1}, x^1_{-m+2}, \ldots$. Let $h_m(x^m)$ be some scalar deterministic operation that is continuous and strictly monotone, with respect to the metric $m^{-1}\sum_{i=1}^m \gamma(x_i, y_i)$. Let v be some random variable whose distribution function $F_v(u)$ is continuous, and such that $F_v(-u) = 1 - F_v(u)$. Then, $F_v(u)$ is also strictly monotone. Given some data sequence x, let us then generate the transformed sequence $\ldots, F_v(h_m(x_{-m}^{-1})),\ F_v(h_m(x^0_{-m+1})),\ F_v(h_m(x^1_{-m+2})), \ldots$. Since the transformation induced by $h_m(\cdot)$ and $F_v(\cdot)$ is stationary, if the data sequence x is generated by some stationary process μ, the induced process $\mu \circ F_v \circ h_m$ is also stationary. Furthermore, due to the continuity and strict monotonicity of both the operations $h_m(\cdot)$ and $F_v(\cdot)$, to estimate the value x_0 of the datum X_0 from the process μ, given either x_{-n}^{-1} or $x_{-n}^{-1} \cup x_1^N$, it suffices to estimate $F_v(h_m(x^0_{-m+1}))$. Finally, due to the boundness of the operation $F_v(\cdot)$, if $\mathscr{M}$ is a compact class of stationary processes, the class $\mathscr{M}(F_v \circ h_m)$ induced by $\mathscr{M}$ and the operations F_v and h_m is also a compact class of stationary processes. Given the class $\mathscr{M}$ of stationary processes and the operations F_v and h_m as above, let us now consider the class of predictors or interpolators,

$$\mathscr{H}_l^3 = \{g_{l,x^l} \colon g_{l,x_{i+1}^{i+l}} = \textstyle\sum_{k=1}^{l+1-m} a_{lk}\, \mathrm{sgn}(h_m(x_{i+k}^{i+k+m-1}) + v_k),$$
$l \geqslant m$,
where $h_m(x^m)$ is continuous and strictly monotone, with respect to the metric $m^{-1}\sum_{i=1}^m \gamma(x_i, y_i)$ scalar and real function, $\{v_k\}$, a sequence of i.i.d. random variables with distribution function $F_v(u)$ that is continuous, and such that $F_v(-u) = 1 - F_v(u)$, $\forall u \in R$, and $\{a_{lk}\}$, some set of real and bounded coefficients$\}$. (K)

The class $\mathscr{H}_l^3$ above clearly satisfies part (i), in Definition 1 of robustness, for l finite. For $l \to \infty$, part (ii) of Definition 1 is satisfied if some additional conditions (such as in (9)) are imposed on the set $\{a_{lk}\}$ of coefficients in $\mathscr{H}_l^3$. Given l, m, $h_m(\cdot)$, $F_v(\cdot)$, let us now evaluate the expected values

$$E\left\{F_v(h_m(X^0_{-m+1})) - \sum_{k=1}^{l+1-m} a_{lk}\,\mathrm{sgn}(h_m(X^{-l+k+m-2}_{-l-1+k}) + v_k)\right\}^2$$

and

$$E\left\{F_v(h_m(X^0_{-m+1})) - \sum_{\substack{1 \leq k \leq l+2-m \\ k \neq -m-i+1}} a_{lk}\,\mathrm{sgn}(h_m(X^{i+k+m-1}_{i+k}) + v_k)\right\}^2, \qquad i < -m. \quad (13)$$

Due to (11), we obtain

$$\begin{aligned} E_\mu&\left\{F_v(h_m(X^0_{-m+1})) - \sum_{k=1}^{l+1-m} a_{lk}\,\mathrm{sgn}(h_m(X^{-l+k+m-2}_{-l-1+k}) + v_k)\right\}^2 \\ &= E_\mu\left\{F_v(X^0_{-m+1})) - \sum_{k=1}^{l+1-m} a_{lk} F_v(h_m(X^{-l+k+m-2}_{-l-1+k}))\right\}^2 \\ &\quad + \left(\sum_{k=1}^{l+1-m} |a_{lk}|^2\right) E_\mu\{F_v(h_m(X^{-l+k+m-2}_{-l-1+k})) - F_v^2(h_m(X^{-l+k+m-2}_{-l-1+k}))\} \end{aligned} \quad (14)$$

$$\begin{aligned} E_\mu&\left\{F_v(h_m(X^0_{-m+1})) - \sum_{\substack{1 \leq k \leq l+2-m \\ k \neq -m-i+1}} a_{lk}\,\mathrm{sgn}(h_m(X^{i+k+m-1}_{i+k}) + v_k)\right\}^2 \quad (1 < -m) \\ &= E_\mu\left\{F_v(h_m(X^0_{-m+1})) - \sum_{\substack{1 \leq k \leq l+2-m \\ k \neq -m-i+1}} a_{lk} F_v(h_m(X^{-l+k+m-2}_{-l-1+k}))\right\}^2 \\ &\quad + \left(\sum_{\substack{1 \leq k \leq l+2-m \\ k \neq -m-i+1}} |a_{lk}|^2\right) E_\mu\{F_v(h_m(X^{-l+k+m-2}_{-l-1+k})) \\ &\quad - F_v^2(h_m(X^{-l+k+m-2}_{-l-1+k}))\}. \end{aligned} \quad (15)$$

If in expressions (14) and (15), we ignore the second terms with the factor $\sum |a_{lk}|^2$, we minimize the first terms with respect to $\{a_{lk}\}$, and the resulting filters are such that $\sum a_{lk} = 1$, $\lim_{l \to \infty} \sum |a_{lk}|^2 \to 0$, then all the conditions (including (9)) for robustness discussed above will be satisfied. The above conditions on the coefficients $\{a_{lk}\}$ will be satisfied if the spectral density of the process $\mu \circ F_v \circ h_m$ has the appropriate form. Consider now the class (K) of operations, and let μ_0 and μ be two stationary processes producing data sequences that are denoted x and y, respectively. Let $\rho(\cdot,\cdot)$ be such that $\rho(u,v) \triangleq (u-v)^2$. Then, we easily obtain the expression

$$\begin{aligned}
&\bar{\rho}(\mu_0 \circ g_l, \mu \circ g_l) \\
&\quad = E_p \left\{ \sum_k a_{lk} [F_v(h_m(X_{i+k}^{i+k+m-1})) - F_v(h_m(Y_{i+k}^{i+k+m-1}))] \right\}^2 \\
&\qquad + \sum_k |a_{lk}|^2 (E_{\mu_0} \{F_v(h_m(X^m))[1 - F_v(h_m(X^m))]\} \\
&\qquad + E_\mu \{F_v(h_m(Y^m))[1 - F_v(h_m(Y^m))]\}),
\end{aligned} \tag{16}$$

where p is the joint stationary measure with marginals μ_0 and μ that yields the rho-bar distance $\bar{\rho}(\mu_0 \circ g_l, \mu \circ g_l)$.

We now express a lemma concerning the sensitivity of the operations in class (K).

LEMMA 6. *Consider the class of operations $\mathscr{H}_l^3$ in* (K). *In addition, let the coefficients $\{a_{lk}\}$ satisfy the conditions* (9), *and let the function $h_m(x^m)$ have a uniformly bounded derivative on R^m with respect to the metric $\gamma_m(x^m, y^m)$. Let the density function of the variables $\{v_k\}$ be uniformly bounded on R. Let $\gamma(\cdot,\cdot)$ be such that $\gamma(x,y) \triangleq |x-y|$, and let $\rho(\cdot,\cdot)$ be such that $\rho(u,v) \triangleq (u-v)^2$. Then, the sensitivity $\mathscr{S}_\rho(\mu_0, \{g_l\})$ (Definition* 5) *of the operations in $\mathscr{H}_l^3$ is bounded at every stationary process μ_0. It follows that the sensitivity $\mathscr{S}_{w\rho}(\mu_0, \{g_l\})$ (Definition* 6) *is also bounded then, at every stationary measure μ_0.*

Proof. Let the density function of the variables $\{v_k\}$ be bounded uniformly by the positive constant c. Let the derivative with respect to $\gamma_m(x^m, y^m)$ of the function $h_m(x^m)$ be uniformly bounded by the positive constant $\mathscr{C}$. Denote $B \triangleq c \cdot \mathscr{C}$. Then,

$$\begin{aligned}
|F_v(h_m(x^m)) - F_v(h_m(y^m))| &\leqslant c\,|h_m(x^m) - h_m(y^m)| \leqslant c \cdot \mathscr{C} \gamma_m(x^m, y^m) \\
&= B\gamma_m(x^m, y^m), \qquad \forall x^m, y^m \in R^m.
\end{aligned} \tag{17}$$

Also,

$$|F_v(h_m(x^m)) - F_v(h_m(y^m))| \leqslant 1, \qquad \forall x^m, y^m \in R^m. \tag{18}$$

Let x and y denote infinite sequences generated respectively by the stationary processes μ_0 and μ. For l given, let $\Pi_{\Pi_{\gamma,m}}(\mu_0^l, \mu^l) = \delta_l$ and let p^l

denote the joint stationary measure that yields the Prohorov distance $\Pi_{\Pi_{\gamma,m}}(\mu_0^l, \mu^l)$. Then,

$$p(x^l, y^l\colon \Pi_{\gamma,m}(\mu_{x^l}^m, \mu_{y^l}^m) > \delta_l) \leqslant \delta_l. \tag{19}$$

Now, from (17), (18), and (19), we obtain

$$\begin{aligned}
&E_{p^l}\left\{\sum_k a_{lk}[F_v(h_m(X_{i+k}^{i+k+m-1})) - F_v(h_m(Y_{i+k}^{i+k+m-1}))]\right\}^2 \\
&\quad\leqslant \left(\sum_k a_{lk}\right)^2 p^l(x^l, y^l\colon \Pi_{\gamma,m}(\mu_{x^l}^m, \mu_{y^l}^m) > \delta_l) \\
&\qquad + \int_{x^l, y^l\colon \Pi_{\gamma,m}(\mu_{x^l}^m, \mu_{y^l}^m) \leqslant \delta_l} \left[\sum_k a_{lk} |F_v(h_m(x_{i+k}^{i+k+m-1})) \right. \\
&\qquad \left. - F_v(h_m(y_{i+k}^{i+k+m-1}))|\right]^2 dp^l(x^l, y^l), \qquad \forall l > l_0.
\end{aligned} \tag{20}$$

But from the definition of the Prohorov distance $\Pi_{\Pi_{\gamma,m}}(\mu_{x^l}^m, \mu_{y^l}^m)$ in (8), we have that $\Pi_{\gamma,m}(\mu_{x^l}^m, \mu_{y^l}^m) \leqslant \delta_l$ implies

$$x^l, y^l\colon \begin{cases} \{\#k\colon \gamma_m(x_{i+k}^{i+k+m-1}, y_{i+k}^{i+k+m-1}) \leqslant \delta_l\} > l(1-\delta_l) \\ \{\#k\colon \gamma_m(x_{i+k}^{i+k+m-1}, y_{i+k}^{i+k+m-1}) > \delta_l\} < l - l(1-\delta_l) = l\delta_l. \end{cases} \tag{21}$$

Applying (17), (18), (19), and (21) to (20), we obtain

$$\begin{aligned}
&E_{p^l}\left\{\sum_k a_{lk}[F_v(h_m(X_{i+k}^{i+k+m-1})) - F_v(h_m(Y_{i+k}^{i+k+m-1}))]\right\}^2 \\
&\quad\leqslant \delta_l\left(\sum_k a_{lk}\right)^2 + \int_{x^l, y^l\colon \Pi_{\gamma,m}(\mu_{x^l}^m, \mu_{y^l}^m) \leqslant \delta_l} \left[B\sum_k a_{lk}\,\delta_l + \sup_{\{i_j\}} \sum_{j=1}^{l\delta_l} a_{i_j}\right]^2 dp^l(x^l, y^l) \\
&\quad= \delta_l\left(\sum_k a_{lk}\right)^2 + \left[B\left(\sum_k a_{lk}\right)\delta_l + \sup_{\{i_j\}} \sum_{j=1}^{l\delta_l} a_{i_j}\right]^2 \\
&\qquad \times p^l(x^l, y^l\colon \Pi_{\gamma,m}(\mu_{x^l}^m, \mu_{y^l}^m) \leqslant \delta_l) \\
&\quad\leqslant \delta_l\left(\sum_k a_{lk}\right)^2 + \left[B\left(\sum_k a_{lk}\right)\delta_l + \sup_{\{i_j\}} \sum_{j=1}^{l\delta_l} a_{i_j}\right]^2, \\
&\qquad\qquad \text{where} \quad \delta_l = \Pi_{\Pi_{\gamma,m}}(\mu_0^l, \mu^l).
\end{aligned} \tag{22}$$

Consider now the condition $\lim_{l\to\infty} \sum_k |a_{lk}|^2 = 0$, on the coefficients $\{a_{lk}\}$. This condition implies that there exists bounded constant β, such that

$$\lim_{l\to\infty} \sup_{\{i_j\}} \sum_{j=1}^{l\delta_l} a_{i_j} < \beta\delta_l. \tag{23}$$

From (16), (22), and (23), and due to the fact that the expectation $E_\mu\{F_v(h_m(X^m)[1-F_v(h_m(X^m)]\}$ is bounded from above by one, for every μ, we obtain

$$\lim_{l\to\infty} \frac{\bar{\rho}(\mu_0 \circ g_l, \mu \circ g_l)}{\Pi_{\Pi_{\gamma,m}}(\mu_0^l, \mu^l)} \leqslant \lim_{l\to\infty} \left(\sum_k a_{lk}\right)^2 + [B+\beta]^2 [\lim_{l\to\infty} \delta_l],$$

$$\text{where} \quad \delta_l = \Pi_{\Pi_{\gamma,m}}(\mu_0^l, \mu^l). \quad (24)$$

And from (24) we thus obtain

$$\lim_{\lim_{l\to\infty} \Pi_{\Pi_{\gamma,m}}(\mu_0^l \mu^l) \to 0} \lim_{l\to\infty} \frac{\bar{\rho}(\mu_0 \circ g_l, \mu \circ g_l)}{\Pi_{\Pi_{\gamma,m}}(\mu_0^l, \mu^l)} \leqslant \lim_{l\to\infty} \left(\sum_k a_{lk}\right)^2 = 1.$$

The proof is now complete.

We will now proceed with another lemma, which exhibits the difference between the sensitivities in Definitions 5 and 6.

LEMMA 7. *Consider linear operations of the form $g_l \triangleq \sum_k a_{lk} x_k$. Let the metric $\gamma(\cdot, \cdot)$ be as in Lemma 6, and let $\rho(u, v) \triangleq \gamma(u, v)$. Then, the sensitivity $\mathscr{S}_{w\rho}(\mu_0, \{g_l\})$ of the operations $\{g_l\}$ is bounded at every measure μ_0, if the limit $\lim_{l\to\infty} (\sum_k |a_{lk}|)$ is bounded.*

Proof. Let μ_0 and μ be two stationary processes, and let x and y denote respectively infinite sequences from μ_0 and μ. Let p be the joint stationary measure that has marginals μ_0 and μ and yields the distance $\bar{\gamma}(\mu_0, \mu)$. Then,

$$\bar{\rho}(\mu_0 \circ g, \mu \circ g) \leqslant E_p \left\{ \left| \sum_k a_k X_k - \sum_k a_k Y_k \right| \right\}$$

$$\leqslant \left(\sum_k |a_{lk}|\right) E_p\{|X_0 - Y_0|\} = \left(\sum_k |a_{lk}|\right) \bar{\gamma}(\mu_0, \mu). \quad (25)$$

Thus, if $\lim_{l\to\infty} (\sum_k |a_{lk}|) \leqslant B$, we obtain

$$\lim_{l\to\infty} \frac{\bar{\rho}(\mu_0 \circ g_l, \mu \circ g_l)}{\bar{\rho}(\mu_0, \mu)} \leqslant B \geqslant \mathscr{S}_{w\rho}(\mu_0, \{g_l\}).$$

The proof of the lemma is now complete.

We note that the operations in Lemma 7 are not even ρ-robust at μ_0. In contrast, not only are they weakly ρ-robust at μ_0, but they also induce bounded sensitivity. We point out here that directly from the derivation and the conclusions in Lemma 6, we also conclude that operations of the form $g_l = \sum_k a_{lk}\, g_m(X_{i+k}^{i+k+m-1})$ are ρ-robust and they induce bounded sensitivity at every process, if the coefficients $\{a_{lk}\}$ satisfy the conditions (9), and the function $g_m(\cdot)$ is deterministic, bounded, and differentiable, with uniformly bounded derivative with respect to the metric $\gamma_m(x^m, y^m)$. Finally, it is easily concluded (as in Boente *et al.* (1982) and Cox (1978)) that if $\gamma(\cdot,\cdot)$ and $\rho(\cdot,\cdot)$ are identical and both are such that $\gamma(u, v) = \rho(u, v) \triangleq (u-v)^2$, then linear operations as in Lemma 7 are weakly ρ-robust and they induce bounded sensitivity at every stationary measure.

We conclude this section with an interesting observation regarding the breakdown point of robust operations. Let $\mathcal{M}$ be a class of stationary processes with common, uniformly bounded spectral density function. Let $\gamma(u, v) = \rho(u, v) \triangleq (u-v)^2$ and consider linear operations as in Lemma 7, where the coefficients $\{a_{lk}\}$ are determined asymptotically $(l \to \infty)$ by the common spectral density function. Then, those operations are weakly ρ-robust on every μ in $\mathcal{M}$, and they induce bounded sensitivity, $\forall \mu \in \mathcal{M}$. However, the asymptotic error that those operations induce at some μ in $\mathcal{M}$ is independent of μ. We thus conclude that the linear operations described above have a breakdown point equal to zero (Definition 6) at every μ in $\mathcal{M}$. The zero breakdown point here means that the linear operations decribed above are totally insensitive to perturbations on measures; they only reflect perturbations in second-order statistics that remain unchanged in $\mathcal{M}$. We note that if the breakdown point had been defined as in Hampel (1971), then it would be infinity in $\mathcal{M}$, and it would not reflect one of the strong weaknesses that characterize the linear operations.

5. The Study of a Special Case

Let us consider predictors and interpolators, in class $\mathcal{H}_l^3$, in (K). Let the sequence $\{v_k\}$, in $\mathcal{H}_l^3$, be a sequence of i.i.d., Gaussian, zero mean, and unit variance random variables. Let $\Phi(x)$ denote the distribution function of a Gaussian random variable as above, let $h_m(x^m)$ be as in the description of the class $\mathcal{H}_l^3$ in (K), and let the prediction of $\Phi(h_m(X_{-m+1}^0))$ be sought. Let the mean squared criterion be used, as a performance measure in prediction, let μ_0 be a nominal, well-known, stationary, real, and scalar stochastic process, and let a compact class, $\mathcal{M}$, of processes be defined by $\mathcal{M} = \{\mu: \mu = (1-\varepsilon)\mu_0 + \varepsilon\nu;\ \nu \in \mathcal{M}_s\}\}$; where $\varepsilon: 0 < \varepsilon < 1$, and $\mathcal{M}_s$ is the class of all real, scalar, zero mean, and unit variance stationary stochastic processes. Let $\mathcal{F}_m$ be the class of real, scalar, and continuous and strictly monotone, with respect to the metric $m^{-1}\sum_{i=1}^m |x_i - y_i|$ functions. Let

$\rho(u, v) \triangleq (u-v)^2$. Then, given $h_m \in \mathscr{F}_m$, $\mu \in \mathscr{M}$, and $\{v_k\}$ as above, we obtain from expression (14) in Section 4,

$$\begin{aligned} e(\mu, h_m, \{a_{lk}\}) & \\ &\triangleq E_\mu \left\{\Phi(h_m(X^0_{-m+1})) - \sum_{k=1}^{l+1-m} a_{lk} \operatorname{sgn}(h_m(X^{-l+k+m-2}_{-l-1+k}) + v_k)\right\}^2 \\ &= E_\mu \left\{\Phi(h_m(X^0_{-m+1})) - \sum_{k=1}^{l+1-m} a_{lk} \Phi(h_m(X^{-l+k+m-2}_{-l-1+k}))\right\}^2 \\ &\quad + \left(\sum_{k=1}^{l+1-m} |a_{lk}|^2\right) E_\mu\{\Phi(h_m(X^m)) - \Phi^2(h_m(X^m))\}. \end{aligned} \tag{26}$$

Let us select $k=1$, and let us denote by $b_{\Phi \cdot h_m \cdot \mu}(\omega)$; $\omega \in [-\pi, \pi]$ the spectral density of the stationary process that is induced by the process μ and the operation $\Phi(h_m(\cdot))$. Let us take $l \to \infty$, and let us denote

$$e(\mu, h_m) \triangleq \inf_{\{a_{l1}\}} \lim_{l \to \infty} e(\mu, h_m, \{a_{l1}\}) \tag{27}$$

$$m(\mu, h_m) \triangleq E_\mu\{\Phi(h_m(X^m))\} \tag{28}$$

$$\sigma^2(\mu, h_m) \triangleq E_\mu\{\Phi(h_m(X^m))\}^2 \tag{29}$$

$$\begin{aligned} f_{\Phi \cdot h_m \cdot \mu}(\omega) \triangleq\ & b_{\Phi \cdot h_m \cdot \mu}(\omega) + m^2(\mu, h_m) \\ & + \left[m(\mu, h_m) - (2\pi)^{-1} \int_{-\pi}^{\pi} b_{\Phi \cdot h_m \cdot \mu}(\omega)\, d\omega - m^2(\mu, h_m)\right] \\ & \times \delta(\omega); \omega \in [-\pi, \pi], \end{aligned} \tag{30}$$

where $\delta(\omega)$ the unit-weight delta function at $\omega = 0$.

Then, due to expressions (27), (28), and (30), and from expression (26), we obtain

$$e(\mu, h_m) = 2\pi \exp\left\{(2\pi)^{-1} \int_{-\pi}^{\pi} \ln f_{\Phi \cdot h_m \cdot \mu}(\omega)\, d\omega\right\}. \tag{31}$$

If $H_{\Phi \cdot h_m \cdot \mu}(\lambda)$; $\lambda \in [-\pi, \pi]$, denotes the Fourier transform of the prediction sequence $\{a_{l1}\}$, that, for $l \to \infty$, yields the minimum error $e(\mu, h_m)$, in (31), then

$$\begin{aligned} & H_{\Phi \cdot h_m \cdot \mu}(\lambda) \colon \|1 - H^*_{\Phi \cdot h_m \cdot \mu}(\lambda)\|^2 \\ & = 2\pi f^{-1}_{\Phi \cdot h_m \cdot \mu}(\lambda) \cdot \exp\left\{(2\pi)^{-1} \int_{-\pi}^{\pi} \ln f_{\Phi \cdot h_m \cdot \mu}(\omega)\, d\omega\right\}, \quad \text{a.e. in } \lambda \in [-\pi, \pi], \end{aligned} \tag{32}$$

where $*$ denotes conjugate.

Let now h_m be given, and let $\mathcal{M}$ be the compact class of processes defined at the beginning of this section. Given μ in $\mathcal{M}$, we easily obtain the following equations from (30),

$$(2\pi)^{-1} \int_{-\pi}^{\pi} f_{\Phi \cdot h_m \cdot \mu}(\omega)\, d\omega = E_\mu\{\Phi(h_m(X^m))\} = m(\mu, h_m) \tag{33}$$

$$\begin{aligned}
&(2\pi)^{-1} \int_{-\pi}^{\pi} b_{\Phi \cdot h_m \cdot \mu}(\omega)\, d\omega \\
&\quad = E_\mu\{\Phi(h_m(X^m))\}^2 - m^2(\mu, h_m) = \sigma^2(\mu, h_m) - m^2(\mu, h_m) \\
&\quad = (1-\varepsilon)[\sigma^2(\mu_0, h_m) - m^2(\mu_0, h_m)] + \varepsilon[\sigma^2(v, h_m) - m^2(v, h_m)] \\
&\qquad + \varepsilon(1-\varepsilon)[m(\mu_0, h_m) - m(v, h_m)]^2
\end{aligned} \tag{34}$$

$$\begin{aligned}
b_{\Phi \cdot h_m \cdot \mu}(\omega) &= (1-\varepsilon)\, b_{\Phi \cdot h_m \cdot \mu_0}(\omega) + \varepsilon b_{\Phi \cdot h_m \cdot v}(\omega) \\
&\quad + \varepsilon(1-\varepsilon)[m(\mu_0, h_m) - m(v, h_m)]^2,
\end{aligned} \tag{35}$$

where $\sigma^2(\mu, h_m)$ is given by expression (29) and $\mu = (1-\varepsilon)\,\mu_0 + \varepsilon v$.

Substituting expression (35) in (30), we obtain

$$\begin{aligned}
f_{\Phi \cdot h_m \cdot \mu}(\omega) &= (1-\varepsilon)\, b_{\Phi \cdot h_m \cdot \mu_0}(\omega) + \varepsilon b_{\Phi \cdot h_m \cdot v}(\omega) \\
&\quad + [(1-\varepsilon)\, m^2(\mu_0, h_m) + \varepsilon m^2(v, h_m)] \\
&\quad + \{(1-\varepsilon)[m(\mu_0, h_m) - \sigma^2(\mu_0, h_m)] \\
&\quad + \varepsilon[m(v, h_m) - \sigma^2(v, h_m)]\}\, \delta(\omega),
\end{aligned} \tag{36}$$

where

$$(2\pi)^{-1} \int_{-\pi}^{\pi} b_{\Phi \cdot h_m \cdot v}(\omega)\, d\omega = \sigma^2(v, h_m) - m^2(v, h_m). \tag{37}$$

Let us now select the following h_m^* function that satisfies the conditions in class $\mathcal{H}_l^3$,

$$h_m^*(X_0^{m-1}) = m^{-1} \sum_{i=0}^{m-1} X_i. \tag{38}$$

We can now express the following lemma.

Lemma 8. *Let μ be some real, scalar, zero-mean, discrete-time stationary process, and let $f_\mu(\omega)$; $\omega \in [-\pi, \pi]$ be its spectral density. Let $h_m^* \circ \mu$ be the stationary process induced by μ, by the function h_m^* in* (38), *and by single-step* *shifting in the sliding block of length m. Let* $f_{h_m^* \circ \mu}(\omega)$, $\omega \in [-\pi, \pi]$, *be the spectral density of the process* $h_m^* \circ \mu$. *Then,*

$$f_{h_m^* \circ \mu}(\omega) = m^{-2} \frac{1 - \cos m\omega}{1 - \cos \omega} \cdot f_\mu(\omega), \qquad \omega \in [-\pi, \pi]. \tag{39}$$

Proof.

$$\begin{aligned}
&E_\mu\{h_m^*(X_0^{m-1})\, h_m^*(X_k^{k+m-1})\} \\
&= m^{-2} \sum_{i=0}^{m-1} \sum_{n=k}^{k+m-1} E_\mu\{X_i X_n\} \\
&= m^{-2} \sum_{i=0}^{m-1} \sum_{n=k}^{k+m-1} (2\pi)^{-1} \int_{-\pi}^{\pi} d\omega\, f_\mu(\omega)\, e^{-j\omega(n-i)} \\
&= m^{-2}(2\pi)^{-1} \int_{-\pi}^{\pi} d\omega\, f_\mu(\omega) \sum_{i=0}^{m-1} \sum_{n=k}^{k+m-1} e^{-j\omega(n-i)} \\
&= m^{-2}(2\pi)^{-1} \int_{-\pi}^{\pi} d\omega\, f_\mu(\omega) \left[\sum_{i=0}^{m-1} e^{j\omega i}\right]\left[\sum_{n=k}^{k+m-1} e^{-j\omega n}\right] \\
&= m^{-2}(2\pi)^{-1} \int_{-\pi}^{\pi} d\omega\, f_\mu(\omega) \cdot \frac{1 - e^{j\omega m}}{1 - e^{j\omega}} \cdot \frac{1 - e^{-j\omega m}}{1 - e^{-j\omega}}\, e^{-j\omega k} \\
&= m^{-2}(2\pi)^{-1} \int_{-\pi}^{\pi} d\omega\, f_\mu(\omega) \frac{1 - \cos m\omega}{1 - \cos \omega}\, e^{-j\omega k} \to f_{h_m^* \circ \mu}(\omega) \\
&= m^{-2} \frac{1 - \cos m\omega}{1 - \cos \omega}\, f_\mu(\omega).
\end{aligned}$$

Let us denote

$$\sigma^2_{h_m^* \circ \mu} \triangleq m^{-2}(2\pi)^{-1} \int_{-\pi}^{\pi} d\omega \frac{1 - \cos m\omega}{1 - \cos \omega} f_\mu(\omega) \tag{40}$$

$$\rho_{h_m^* \circ \mu}(k) \triangleq m^{-2}(2\pi)^{-1} \int_{-\pi}^{\pi} d\omega \frac{1 - \cos m\omega}{1 - \cos \omega} f_\mu(\omega)\, e^{-j\omega k} \tag{41}$$

$$r_{\Phi \circ h_m^* \circ \mu}(k) = (2\pi)^{-1} \int_{-\pi}^{\pi} d\omega\, b_{\Phi \circ h_m^* \circ \mu}(\omega)\, e^{-j\omega k}. \tag{42}$$

Then, we can express the following lemma, whose proof is in the Appendix.

LEMMA 9. *Let* $\mathcal{M}_s$ *be the class of real, scalar, zero-mean, unit variance, discrete-time stationary processes. Given* μ *in* $\mathcal{M}_s$, *given* h_m^* *as in* (38), *let the process* $h_m^* \circ \mu$ *be as in Lemma* 9. *Let* μ_G *be any Gaussian process in* $\mathcal{M}_s$. *Then,*

$$m(\mu_G, h_m^*) = \frac{1}{2} \tag{43}$$

$$r_{\Phi \circ h_m^* \circ \mu_G}(k) = (2\pi)^{-1} \sin^{-1}\left(\frac{\rho_{h_m^* \circ \mu_G}(k)}{1 + \sigma^2_{h_m^* \circ \mu_G}}\right), \qquad k \neq 0 \tag{44}$$

$$r_{\Phi \circ h_m^* \circ \mu_G}(0) = \pi^{-1} \tan^{-1}([2\sigma^2_{h_m^* \circ \mu_G} + 1]^{1/2}) - \frac{1}{4}. \tag{45}$$

Let $\mathcal{M}_s$ be the same class of stationary processes, as in Lemma 9. Given h_m^* as in (38), given μ in $\mathcal{M}_s$, the spectral density $b_{\Phi \circ h_m^* \circ \mu}(\omega)$, $\omega \in [-\pi, \pi]$, will attain its "flattest" form for memoryless measures μ. It will attain its most concentrated form for deterministic measures μ. Let μ_{d} denote some deterministic measure in $\mathcal{M}_s$. Then, we easily obtain

$$b_{\Phi \circ h_m^* \circ \mu_{\mathrm{d}}}(\omega) = [\sigma^2(\mu_{\mathrm{d}}, h_m^*) - m^2(\mu_{\mathrm{d}}, h_m^*)]\, \delta(\omega), \tag{46}$$

where $m(\mu_{\mathrm{d}}, h_m^*)$ and $\sigma^2(\mu_{\mathrm{d}}, h_m^*)$ are given respectively by (28) and (29).

We now express the following lemma, whose proof is in the Appendix.

LEMMA 10. *Let $\mathcal{M}_s$ be as in Lemma 9. Let h_m^* be as in* (38), *and let $m \to \infty$. Then, the two extreme spectral densities $b_{\Phi \circ h_m^* \circ \mu}(\omega)$, $\omega \in [-\pi, \pi]$,* for $\mu \in \mathcal{M}_s$, are given by

$$b_{\Phi \circ h_m^* \circ \mu^*}(\omega) = \frac{1}{12}\, \delta(\omega) \tag{47}$$

$$b_{\Phi \circ h_m^* \circ \nu^*}(\omega) = \frac{(2\pi)^{-1}}{m(m+1)} \frac{1 - \cos m\omega}{1 - \cos \omega}, \qquad \omega \in [-\pi, \pi] \tag{48}$$

$$m(\nu^*, h_m^*) = \frac{1}{2}, \tag{49}$$

where μ^* is the deterministic Gaussian process in $\mathcal{M}_s$ and where ν^* is any memoryless process in $\mathcal{M}_s$.

It can be seen easily that as m increases to asymptotically large values, the spectral density $b_{\Phi \circ h_m^* \circ \nu^*}(\omega)$ tends to $(2\pi)^{-1}\, \delta(\omega)$. Therefore, for h_m^* as in (38), for zero-mean and unit-variance nominal process μ_0 in $\mathcal{M}$, and due to Lemma 10 and expression (33), we have from (36)

$$\begin{aligned} \lim_{m \to \infty} f_{\Phi \circ h_m^* \circ \mu}(\omega) = {} & (1-\varepsilon)\, m^2(\mu_0, h_m^*) + \varepsilon m^2(\nu, h_m^*) \\ & + (2\pi)\{(1-\varepsilon)\, m(\mu_0, h_m^*)[1 - m(\mu_0, h_m^*)] \\ & + \varepsilon m(\nu, h_m^*)[1 - m(\nu, h_m^*)]\}\, \delta(\omega), \\ & \forall \mu \in \mathcal{M},\ \mu = (1-\varepsilon)\, \mu_0 + \varepsilon \nu. \end{aligned} \tag{50}$$

In addition, if the class $\mathscr{M}$ contains zero-mean, unit-variance measures that are also ergodic, then $\lim_{m\to\infty} m(\mu, h_m^*) = \int \Phi(u)\,\delta(u) = \frac{1}{2}$, $\forall \mu \in \mathscr{M}$. So then,

$$\lim_{m\to\infty} f_{\Phi \circ h_m^* \circ \mu}(\omega) = \frac{1}{4} + \frac{\pi}{2}\,\delta(\omega) \triangleq f_{\mathscr{M}}(\omega), \qquad \forall \mu \in \mathscr{M}, \tag{51}$$

$$\int_{-\pi}^{\pi} \ln f_{\mathscr{M}}(\omega)\, d\omega = (2\pi+1)\ln 2^{-2} + \ln(1+2\pi) = \ln(2\pi+1)\cdot 2^{-2(2\pi+1)}. \tag{52}$$

Due to (51) and (52), and from (31) and (32), we also obtain

$$\begin{aligned}\lim_{m\to\infty} e(\mu, h_m^*) &= 2\pi \exp\{[1+(2\pi)^{-1}]\ln 2^{-2} + (2\pi)^{-1}\ln(1+2\pi)\}\\ &= 2\pi(2\pi+1)^{1/2\pi}\, 2^{-[2+\pi^{-1}]}, \qquad \forall \mu \in \mathscr{M},\end{aligned} \tag{53}$$

$$\begin{aligned}&\lim_{m\to\infty} H_{\Phi \circ h_m^* \circ \mu}(\lambda)\colon \|1 - \lim_{m\to\infty} H_{\Phi \circ h_m^* \circ \mu}(\lambda)\|^2\\ &\quad = 8\pi(2\pi+1)^{1/2\pi}\, 2^{-[2+\pi^{-1}]}, \qquad \text{a.e. in } \lambda \in [-\pi, \pi],\ \forall \mu \in \mathscr{M}.\end{aligned} \tag{54}$$

From expressions (53) and (54) we conclude that for h_m^* as in (38), for class $\mathscr{M}$ of zero-mean, unit-variance, ergodic processes, and for $m \to \infty$, at any μ in $\mathscr{M}$, the optimal linear prediction operation in class $\mathscr{H}_l^3$ is given by (54). This operation clearly satisfies conditions (9); thus, it is robust at all μ in $\mathscr{M}$. At all μ in $\mathscr{M}$, the above operation induces error equal to the expression in (53). Therefore, within the above class $\mathscr{M}$, the breakdown point of the overall asymptotic robust operation is one. Finally, the overall operation induces bounded sensitivity everywhere in $\mathscr{M}$. The robust operation studied induces the same characteristics in interpolation as well.

APPENDIX

Proof of Lemma 9. (1)

$$\begin{aligned} m(\mu_G, h_m^*) &= \frac{1}{\sigma_{h_m^* \circ \mu_G}} \int_{-\infty}^{\infty} \Phi(u)\,\phi\left(\frac{u}{\sigma_{h_m^* \circ \mu_G}}\right) du\\ &= \int_{-\infty}^{\infty} \Phi(u\sigma_{h_m^* \circ \mu_G})\,\phi(u)\, du = \int_0^{\infty} \Phi(u\sigma_{h_m^* \circ \mu_G})\,\phi(u)\, du\\ &\quad + \int_{-\infty}^{0} [1 - \Phi(-u\sigma_{h_m^* \circ \mu_G})]\,\phi(-u)\, du\\ &= \int_0^{\infty} \Phi(u\sigma_{h_m^* \circ \mu_G})\,\phi(u)\, du + \frac{1}{2} - \int_0^{\infty} \Phi(u\sigma_{h_m^* \circ \mu_G})\,\phi(u)\, du = \frac{1}{2}.\end{aligned}$$

So,

$$m(\mu_G, h_m^*) = \tfrac{1}{2}. \tag{A.1}$$

For simplicity in notation, let us denote $\sigma \triangleq \sigma_{h_m^* \cdot \mu_G}$ and $\rho_k \triangleq \rho_{h_m^* \circ \mu_G}(k)$. Then, we obtain

$$\begin{aligned} &r_{\Phi \cdot h_m^* \cdot \mu_G}(k) + m^2(\mu_G, h_m^*) \\ &= \iint du\, dv\, \Phi(u)\, \Phi(v) \frac{1}{\sigma} \phi\left(\frac{u}{\sigma}\right) \frac{\sigma}{\sqrt{\sigma^4 - \rho_k^2}} \phi\left(\frac{\sigma}{\sqrt{\sigma^4 - \rho_k^2}}\left[v - \frac{\rho_k}{\sigma^2} u\right]\right) \\ &= \frac{1}{\sigma} \int du\, \Phi(u)\, \Phi\left(\frac{\rho_k}{\sigma\sqrt{\sigma^4 + \sigma^2 - \rho_k^2}} u\right) \phi\left(\frac{u}{\sigma}\right) \\ &= \int du\, \Phi(\sigma u)\, \Phi\left(\frac{r_k \sigma}{\sqrt{1 + \sigma^2(1 - r_k^2)}}\right) \phi(u) \triangleq g(\sigma), \end{aligned} \tag{A.2}$$

where

$$r_k \triangleq \rho_k / \sigma^2 < 1, \qquad \forall k, \tag{A.3}$$

$$g(0) = \tfrac{1}{4}. \tag{A.4}$$

From (A.2) we have, after some transformations,

$$\frac{\partial}{\partial \sigma} g(\sigma) = \frac{r_k \sigma}{\pi(1 + \sigma^2)[1 + 2\sigma^2 + \sigma^4(1 - r_k^2)]^{1/2}}. \tag{A.5}$$

Then, from (A.4) and (A.5), we obtain

$$\begin{aligned} g(\sigma) - \frac{1}{4} &= \int_0^\sigma \frac{r_k}{\pi} \frac{x\, dx}{(1 + x^2)[(1 + x^2)^2 - x^4 r_k^2]^{1/2}} \\ &= \frac{1}{2\pi} \int_0^{r_k \sigma^2 (1 + \sigma^2)^{-2}} [1 - w^2]^{-1/2}\, dw = \frac{1}{2\pi} \int_0^{\sin^{-1}[r_k \sigma^2 (1 + \sigma^2)^{-1}]} dy \\ &= \frac{1}{2\pi} \sin^{-1}[r_k \sigma^2 (1 + \sigma^2)^{-1}] = \frac{1}{2\pi} \sin^{-1}[\rho_k (1 + \sigma^2)^{-1}]. \end{aligned} \tag{A.6}$$

Finally, from (A.1), (A.3), and (A.6), we obtain Eq. (44).

(2)

$$r_{\Phi \cdot h_m^* \circ \mu_G}(0) + \frac{1}{4} = \frac{1}{\sigma} \int \Phi^2(x)\, \phi\left(\frac{x}{\sigma}\right) dx = \int \Phi^2(\sigma x)\, \phi(x)\, dx \triangleq h(\sigma),$$

$$\text{where} \quad h(1) = \frac{1}{3} \tag{A.7}$$

$$\frac{\partial h(\sigma)}{\partial \sigma}=2\int \Phi(\sigma x)\,\phi(\sigma x)\,x\phi(x)\,dx=2(2\pi)^{-1/2}\int \Phi(\sigma x)\,x\phi(x[1+\sigma^2]^{1/2})\,dx$$

$$=-2(2\pi)^{-1/2}(1+\sigma^2)^{-1}\int \Phi(\sigma x)\,\phi'(x[1+\sigma^2]^{1/2})\,dx$$

$$=2\sigma(2\pi)^{-1/2}(1+\sigma^2)^{-1}\int \phi(\sigma x)\,\phi(x[1+\sigma^2]^{1/2})\,dx$$

$$=2\sigma(2\pi)^{-1}(1+\sigma^2)^{-1}\int \phi(x[2\sigma^2+1]^{1/2})\,dx$$

$$=\frac{\sigma}{\pi(1+\sigma^2)[2\sigma^2+1]^{1/2}}. \tag{A.8}$$

So,

$$h(\sigma)-\frac{1}{3}=\frac{1}{\pi}\int_1^{\sigma}\frac{y\,dy}{(1+y^2)[2y^2+1]^{1/2}}$$

$$=\frac{1}{\pi}\int_{3^{1/2}}^{[2\sigma^2+1]^{1/2}}\frac{dx}{x^2+1}=\frac{1}{\pi}\int_{\tan^{-1}(3^{1/2})}^{\tan^{-1}([2\sigma^2+1]^{1/2})}dw$$

$$=\frac{1}{\pi}\tan^{-1}([2\sigma^2+1]^{1/2})-\frac{1}{3}. \tag{A.9}$$

From (A.7) and (A.9) we obtain expression (45).

Proof of Lemma 10. (1) The spectral density for deterministic processes is given by (46). It remains to maximize the quantity $[\sigma^2(\mu_d, h_m^*)-m^2(\mu_d, h_m^*)]$ among all the deterministic processes. But, $f_{\mu_d}(\omega)=\delta(\omega)$ and

$$m^2\left.\frac{1-\cos m\omega}{1-\cos \omega}\right|_{\omega=0}=1.$$

Thus, $f_{h_m^*\cdot\mu_d}(\omega)=\delta(\omega)$, where $f_{h_m^*\cdot\mu}(\omega)$ is given by (39).

Let v be some zero-mean, unit variance, absolutely continuous random variable. Let F_v and f_v denote respectively the distribution and the density functions of the variable v. Let $\mathscr{F}$ be the class of real and monotonically increasing functions, that take values on [0, 1]. Then, it is well known that

$$\sup_{g\in\mathscr{F}}\left\{\int g^2(u)\,f_v(u)\,du-\left[\int g(u)f_v(u)\,du\right]^2\right\}$$

$$=\int F_v^2(u)\,f_v(u)\,du-\left[\int F_v(u)\,f_v(u)\,du\right]^2=\frac{1}{12} \tag{A.10}$$

If ν^* in $\mathscr{M}_s$ is the memoryless Gaussian process, and h_m^* is as in (38), then

$$\sigma^2(\nu^*, h_m^*) - m^2(\nu^*, h_m^*) = \int \Phi^2(u)\,\phi(u)\,du - \left[\int \Phi(u)\,\phi(u)\,du\right]^2 = \frac{1}{12},$$

due to (A.10), where $\phi(u)$ is the unit variance, zero-mean, Gaussian density. The proof of (47) is now complete.

(2) Let ν^* be some memoryless process in $\mathscr{M}_s$. Then, $f_{\nu^*}(\omega) = 1/2\pi$, $\omega \in [-\pi, \pi]$. From (41) we then obtain

$$\rho_{h_m^*, \nu^*}(k) = \begin{cases} (m-k)m^{-2}, & 0 \leqslant k \leqslant m-1 \\ 0, & k \geqslant m. \end{cases} \tag{A.11}$$

For $m \to \infty$, and due to the law of large numbers, the variables $h_m^*(X_0^m)$ and $h_m^*(X_k^{k+m})$ are jointly Gaussian for all k. Then, substituting (A.11) in (44) and (45), we obtain

$$r_{\Phi, h_m^*, \nu^*}(0) = \pi^{-1} \tan^{-1}([2m^{-1}+1]^{1/2}) - \frac{1}{4}$$
$$r_{\Phi, h_m^*, \nu^*}(k) = \begin{cases} (2\pi)^{-1} \sin^{-1}\left(\dfrac{m-k}{m(m+1)}\right), & 1 \leqslant k \leqslant m-1, \\ 0, & k \geqslant m. \end{cases} \tag{A.12}$$

For large m we write

$$\begin{gathered} \sin^{-1}\left(\frac{m-k}{m(m+1)}\right) \sim \frac{m-k}{m(m+1)}, \qquad 1 \leqslant k \leqslant m-1, \\ r_{\Phi, h_m^*, \nu^*}(0) \sim (2\pi)^{-1} \frac{1}{m+1}. \end{gathered} \tag{A.13}$$

From (A.12) and (A.13), we obtain (48). We also obtain (49), due to (43).

RECEIVED September 26, 1984; ACCEPTED October 3, 1986

REFERENCES

BOENTE, G., FRAIMAN, R., AND YOHAI, V. (1982), "Qualitative Robustness for General Stochastic Processes," Department of Statistics, University of Washington Seattle, Technical Report No. 26, Oct.; submitted for publication.

COX, D. (1978), "Metrics on Stochastic Processes and Qualitative Robustness," Department of Statistics, University of Washington Seattle, Technical Report No. 23.

HAMPEL F. R. (1971), A general qualitative definition of robustness, *Ann. Math. Statist.* **42**, 1887–1896.

HOSOYA, Y. (1978), Robust linear extrapolation of second order stationary processes, *Ann. Probab.* **6**, 574–584.

KOLMOGOROV, A. (1941), Interpolation and extrapolation," *Bull. Acad. Sci. USSR, Ser. Math.* **5**, 3–14.

MARTIN, R. D. AND DEBOW, G. (1976), Robust filtering with data dependent covariance, *in* "Proceedings, Johns Hopkins Conf. on Information Sciences and Systems."

MASRELIEZ C. J., AND MARTIN, R. D. (1977), Robust Bayesian estimation for the linear model and robustifying the Kalman filter, *IEEE Trans. Automat. Control* **AC-22**, June, 361–371.

PAPANTONI-KAZAKOS P. (1981a), Sliding block encoders that are rho-bar continuous functions of their input, *IEEE Trans. Inform. Theory* **IT-27** May, 372–376.

PAPANTONI-KAZAKOS, P. (1981b), Stochastic quantization for performance stability, *Inform. and Control* **49**, 171–198.

PAPANTONI-KAZAKOS, P. (1982), "Performance Bounds in Robust Filtering and Smoothing," University of Connecticut, EECS Department, Technical Report TR-82-6, Oct.

PAPANTONI-KAZAKOS, P. (1984) A game theoretic approach to robust filtering, *Inform. and Control* **60**, 1735–1758.

PAPANTONI-KAZAKOS, P., AND GRAY, R. M. (1979), Robustness of estimators on stationary observations, *Ann. Probab.* **7**, 989–1002.

TSAKNAKIS, H., KAZAKOS, D., AND PAPANTONI-KAZAKOS, P. (1982), "Robust Prediction and Interpolation for Vector Stationary Processes," University of Connecticut, EECS Department, Technical Report TR-82-7, Nov.; submitted for publication.

TSAKNAKIS, H., AND PAPANTONI-KAZAKOS, P. (1983), "Robust Linear Filtering for Multivariable Stationary Time Series," University of Connecticut, EECS Departement Technical Report TR-83-6, April; submitted for publication.

WIENER, N. (1949), Extrapolation, "Interpolation, and Smoothing of Stationary Time Series, MIT Press, Cambridge, Ma.